Enhancing Scientific Reproducibility through Transparent Reporting

September 25–26, 2019

National Academy of Sciences Building, Lecture Room
2101 Constitution Ave. NW, Washington, DC 20418
Washington, DC

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Enhancing Reproducibility through Transparent Reporting – A Workshop

September 25-26, 2019

Briefing Book Table of Contents

Meeting Materials

Background Information

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Enhancing Scientific Reproducibility through Transparent Reporting

A Workshop

September 25–26 2019 ▪ Washington, DC

Statement of Task

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine (National Academies) will plan and conduct a public workshop to discuss the current state of transparency in reporting biomedical research (e.g. disclosure of the availability and location of data, materials, analysis, and methodology) and to explore the possibility of improving the harmonization of guidelines across journals and funding agencies so that biomedical researchers propose and report data in a consistent manner. This workshop is sponsored by the National Institutes of Health, Cell Press, *The Lancet*, and Nature Research.

Workshop objectives:

- Highlight current efforts by researchers, institutions, funders, and journals to increase transparency in proposing and reporting pre-clinical biomedical research;
- Discuss journal and funder assessments of researchers' adherence to reporting guidelines, including a discussion of the effectiveness of checklists;
- Consider lessons learned from field-specific best practices for increased transparency in reporting rigor elements (i.e., research design, methodology, analysis, interpretation and reporting of results) that are generalizable across biomedical research domains;
- Discuss opportunities for improving the consistency of reporting guidelines and requirements for rigor and transparency by journals, funders, and institutions across the biomedical research lifecycle; and
- Consider approaches to compare reporting of rigor elements proposed in grant applications to those included in publications.

The committee will plan and organize the workshop, develop the agenda, select and invite speakers and discussants, and moderate or identify moderators for the discussions. The agenda will include a panel discussion on facilitating the development of consistent guidelines (e.g. a common set of minimal reporting standards) that could be applied across journals and funders to increase transparency in proposing and reporting biomedical research.

A proceedings of the presentations and discussions at the workshop will be prepared by a designated rapporteur in accordance with institutional guidelines.

Planning Committee

Harvey Fineberg (Chair), Gordon and Betty Moore Foundation

Otis Brawley, Johns Hopkins University

Barry Coller, The Rockefeller University

Stuart Hoffman, U.S. Department of Veterans Affairs

Veronique Kiermer, *PLOS*

Benedict Kolber, Duquesne University

Alexa McCray, Harvard Medical School

Jill Mesirov, UC San Diego School of Medicine

Martin Murphy, CEO Roundtable on Cancer

Richard Nakamura, (Formerly) Center for Scientific Review, NIH

Franklin Sayre, University of Minnesota

Ida Sim, University of California, San Francisco

Valda Vinson, American Association for the Advancement of Science

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Enhancing Scientific Reproducibility through Transparent Reporting

A Workshop

September 25 – 26, 2019
National Academy of Sciences Building, Lecture Room
2101 Constitution Ave. NW, Washington, DC 20418

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine is convening a public workshop to discuss the current state of transparency in reporting pre-clinical biomedical research (e.g., disclosure of the availability and location of data, materials, analysis, and methodology) and to explore the possibility of improving the harmonization of guidelines across journals and funding agencies so that biomedical researchers propose and report data in a consistent manner. This workshop is sponsored by the National Institutes of Health, Cell Press, *The Lancet*, and Nature Research.

WORKSHOP OBJECTIVES:

- Highlight current efforts by researchers, institutions, funders, and journals to increase transparency in proposing and reporting pre-clinical biomedical research;
- Consider lessons learned from field-specific best practices for increased transparency in reporting rigor elements (i.e., research design, methodology, analysis, interpretation and reporting of results) that are generalizable across biomedical research domains;
- Discuss journal and funder assessments of researchers' adherence to transparent reporting guidelines, including a discussion of the effectiveness of checklists;
- Discuss opportunities for improving the consistency of reporting guidelines and requirements for rigor and transparency by journals, funders, and institutions across the biomedical research lifecycle; and
- Consider approaches to compare reporting of rigor elements proposed in grant applications to those included in publications.

DAY 1: September 25, 2019

8:00 a.m. Breakfast available outside the Lecture Room

8:30 a.m. **Welcome and opening remarks**
HARVEY FINEBERG, *Workshop Chair*
President
Gordon and Betty Moore Foundation

Highlights and related recommendations from the National Academies report on Reproducibility and Replicability in Science

9:15 a.m. *Q&A with audience*

SESSION I CULTIVATING TRANSPARENT REPORTING IN BIOMEDICAL RESEARCH

Session Objectives:

- Highlight current efforts by researchers, institutions, funders, and journals to increase transparency in proposing and reporting pre-clinical biomedical research
- Discuss the incentives, disincentives, challenges, and opportunities for researchers when it comes to transparent reporting of pre-clinical biomedical research (e.g., pressure to publish, institutional resources, training, funding).
- Discuss experience with implementation of policies to encourage transparent reporting across the biomedical research life cycle.
- Consider the role of stakeholders in supporting a cultural shift towards transparent reporting in preclinical biomedical research.

  For more information on cultural barriers as sources of non-reproducibility, see p. 58, p. 97, and p. 104 of the National Academies' Reproducibility and Replicability in Science report.

9:30 a.m. ***Opening remarks by session moderator***
ALEXA MCCRAY
Professor of Medicine
Harvard Medical School

9:40 a.m. ***A researcher (early career) perspective***
YARIMAR CARRASQUILLO
Investigator
National Center for Complementary and Integrative Health, National Institutes of Health

9:55 a.m. ***A researcher/researcher support perspective***
BRIAN NOSEK
Co-founder
Center for Open Science

10:10 a.m. ***A researcher (later career)/society publisher perspective***
ARTURO CASADEVALL
Professor, Molecular Microbiology and Immunology, Johns Hopkins University
Editor-in-chief, mBio

10:25 a.m. ***An NIH perspective***
CARRIE WOLINETZ
Acting Chief of Staff and Associate Director for Science Policy
Office of the Director, National Institutes of Health (NIH)

10:40 a.m. *Audience Q&A with the panel*

Discussion Questions:

- *What forces are influencing the culture of biomedical research, and how is it changing?*
- *What actions could influence practice and support a cultural shift towards more transparent reporting?*
- *What influence might transparent reporting or required reporting of rigor elements have on grant applications? Is there a role for pre-registration of pre-clinical studies?*

11:10 a.m. **BREAK**

SESSION II ANSWERING THE CALL FOR CHANGE: LESSONS LEARNED AND BEST PRACTICES

Session Objectives:

- Consider lessons learned from institutional and/or field-specific best practices for increased transparency in reporting rigor elements (i.e. research design, methodology, analysis, interpretation and reporting of results) that are generalizable across biomedical research domains.
- Consider available tools and best practices for increased transparent reporting that support researchers and are generalizable across biomedical research domains.
- Discuss the roles of educational institutions, professional societies, researchers, and funders in improving computational reproducibility (*Reproducibility and Replicability in Science* Report Recommendation 6-6).
- Discuss how funding agencies and organizations could invest in research and development of open-source, usable tools and infrastructure that support reproducibility for a broad range of studies across different domains in a seamless fashion, as well as in outreach to inform and train researchers on best practices (*Reproducibility and Replicability in Science* Report Recommendation 6-1).

11:30 a.m. ***Opening remarks by session moderator***
VERONIQUE KIERMER
Executive Editor
PLOS

11:40 a.m. ***A clinical researcher perspective: Lessons from the SPIRIT initiative***
AN-WEN CHAN
Phelan Scientist, Women's College Research Institute
Associate Professor, University of Toronto

11:50 a.m. ***An institution perspective***
GEETA SWAMY
Vice Dean for Scientific Integrity
Associate Vice President for Research
Duke University

12:00 p.m.. ***A funder perspective***
MAGALI HAAS
Chief Executive Officer and President
Cohen Veterans Bioscience

12:10 p.m. *Moderated panel discussion among speakers*

12:30 p.m. *Audience Q&A with the panel*

Discussion Questions:

- *How can challenges with pre-registration, image analysis, cell line authentication, statistical analysis, or other rigor elements be addressed?*
- *What actions can institutions or professional societies take to educate and support their constituents on best practices? How could this information be best provided?*
- *How might funding agencies and organizations invest in development of open-source reusable tools and infrastructure to support transparent reporting seamlessly across different domains?*
- *What actions could funding agencies and organizations take to inform, train, and support researchers on best practices in transparent reporting?*
- *What has been learned from open access mandates and from implementing policies around sharing data in preclinical research? How could those lessons inform transparent reporting guidance and adoption?*

1:00 p.m. **BREAK** (Lunch available Outside the Lecture Room)

SESSION III STAKEHOLDER PERSPECTIVES ON CHECKLISTS AND GUIDELINES

Session Objectives:

- Discuss journal and funder assessments of researchers' adherence to transparent reporting guidelines, including discussion of the effectiveness of checklists.
  - Highlight empirical assessments of checklist application from funders, journals, and researchers; and
  - Consider practical application and effectiveness of checklists and guidelines to encourage or require transparent reporting of pre-clinical biomedical research.
- Discuss how funders could require thoughtful discussion in grant applications of how uncertainties will be evaluated, along with any relevant issues regarding replicability and computational reproducibility (*Reproducibility and Replicability in Science* Report Recommendation 6-9)
- Discuss how journals and scientific societies could disclose their policies relevant to achieving reproducibility and replicability; and how journals could be encouraged to set and implement desired standards of reproducibility and replicability and adopt policies to reduce the likelihood of non-replicability (*Reproducibility and Replicability in Science* Report Recommendation 6-7)

2:00 p.m. ***Opening remarks by session moderator***
BARRY COLLER
Physician-in-Chief, Vice President for Medical Affairs, and David Rockefeller Professor
The Rockefeller University

2:10 p.m. ***The checklist approach at life science journals – challenges and opportunities***
SOWMYA SWAMINATHAN
Head of Editorial Policy
Nature Research

MALCOLM MACLEOD
Professor
University of Edinburgh

2:30 p.m. ***An NIH funder perspective***
SHAI SILBERBERG
Director Research Quality
National Institute of Neurological Disorders and Stroke, National Institutes of Health

2:40 p.m. *Moderated panel discussion among speakers*

3:10 p.m. *Audience Q&A with the panel*

Discussion Questions:

- *How valuable are checklists? How valuable is guidance such as the CONSORT statement? What are observed challenges to adherence, and how could they be addressed?*
- *How could checklists be improved and/or complemented to further encourage transparent reporting?*
- *What resources do researchers need to be able to submit proposals, publish, or otherwise report on specific rigor elements?*
- *How might funders require thoughtful discussion in grant applications of how uncertainties (such as in measurement, computation, knowledge, modeling, or methods of analysis) will be evaluated by researchers?*
- *Should scientific societies be encouraged to develop policies relevant to transparent reporting?*

SESSION IV PART 1 TOWARDS MINIMAL REPORTING STANDARDS FOR PRECLINICAL BIOMEDICAL RESEARCH

Session Objectives:

- Discuss opportunities for improving the consistency of reporting guidelines and requirements for rigor and transparency by journals, funders, and institutions across the biomedical research lifecycle.

4:00 p.m. **Discussion with audience on potential steps stakeholders could take to support harmonizing reporting guidelines**

HARVEY FINEBERG, *Workshop Chair and session moderator*
President
Gordon and Betty Moore Foundation

BENEDICT KOLBER
Associate Professor
Duquesne University

RICHARD NAKAMURA
Former Director (Retired)
Center for Scientific Review, National Institutes of Health

FRANKLIN SAYRE
STEM Librarian
Thompson Rivers University

VALDA VINSON
Editor, Research
Science

5:00 p.m. **ADJOURN WORKSHOP DAY 1**

DAY 2: September 26, 2019

8:00 a.m. Breakfast Available Outside the Lecture Room

8:30 a.m. **Welcome and overview of Day 1**
HARVEY FINEBERG, *Workshop Chair*
President
Gordon and Betty Moore Foundation

9:00 a.m. **Keynote Address**
MARCIA MCNUTT
President
National Academy of Sciences

9:20 a.m. **Q&A Session**

9:30 a.m. **BREAK**

SESSION IV PART 2 TOWARDS MINIMAL REPORTING STANDARDS FOR PRECLINICAL BIOMEDICAL RESEARCH

Session Objectives:

- Consider approaches to compare reporting of rigor elements proposed in grant applications to those included in publications.
- Suggest stakeholder actions to encourage transparent reporting and practical next steps towards establishing minimal reporting standards for pre-clinical biomedical research.

10:00 a.m. ***Opening remarks by session moderator***
HARVEY FINEBERG, *Workshop Chair*
President
Gordon and Betty Moore Foundation

10:10 a.m. ***An early career researcher perspective***
MICHAEL KEISER
Assistant Professor
University of California, San Francisco

10:20 a.m. ***An institution perspective***
MELISSA RETHLEFSEN
Associate Dean, George A. Smathers Libraries
Fackler Director, Health Science Center Libraries
University of Florida

10:30 a.m. ***A research educator perspective***
STEVEN GOODMAN
Professor of Medicine and Health Research and Policy
Co-director, Meta-Research Innovation Center at Stanford
Stanford University

10:40 a.m. *Moderated panel discussion among speakers*

11:10 a.m. *Small group table discussion and reporting*

Discussion Questions:

- *What actions should funders, researchers, institutions, and journals take to drive widespread adoption of minimal reporting standards?*
- *Are reporting categories in guidelines for publishing (such as materials, design, analysis, and reporting) relevant for funders? For institutions? For small publishers/professional societies?*
- *What other information or reporting categories would be relevant?*
- *How should funders instruct reviewers of grant applications to reinforce transparent reporting? How much information should funders request, that is, to what level of detail, in grant applications)? Is it possible to obtain sufficient information about transparent reporting in grant applications without dramatic expansion of the application?*

12:25 p.m. **Workshop wrap up and concluding discussion with audience**

12:30 p.m. **ADJOURN WORKSHOP DAY 2**

ABOUT THE FORUM

The Forum on Drug Discovery, Development, and Translation of the National Academies of Sciences, Engineering, and Medicine was created in 2005 by the Board on Health Sciences Policy to provide a unique platform for dialogue and collaboration among thought leaders and stakeholders in government, academia, industry, foundations, and patient advocacy with an interest in improving the system of drug discovery, development, and translation. The Forum brings together leaders from private sector sponsors of biomedical and clinical research, federal agencies sponsoring and regulating biomedical and clinical research, the academic community, and patients, and in doing so serves to educate the policy community about issues where science and policy intersect. The Forum convenes several times each year to identify, discuss, and act on key problems and strategies in the discovery, development, and translation of drugs. To supplement the perspectives and expertise of its members, the Forum also holds public workshops to engage a wide range of experts, members of the public, and the policy community. The Forum also fosters collaborations among its members and constituencies. The activities of the Forum are determined by its members, focusing on the major themes outlined below.

INNOVATION AND THE DRUG DEVELOPMENT ENTERPRISE

Despite exciting scientific advances, the pathway from basic science to new therapeutics faces challenges on many fronts. New paradigms for discovering and developing drugs are being sought to bridge the ever-widening gap between scientific discoveries and translation of those discoveries into life-changing medications. There is also increasing recognition of the need for new models and methods for drug development and translational science, and "precompetitive collaborations" and other partnerships, including public–private partnerships, are proliferating. The Forum offers a venue to discuss effective collaboration in the drug discovery and development enterprise and also hosts discussions that could help chart a course through the turbulent forces of disruptive innovation in the drug discovery and development "ecosystem."

Key gaps remain in our knowledge about science, technology, and methods needed to support drug discovery and development. Recent rapid advances in innovative drug development science present opportunity for revolutionary developments of new scientific techniques, therapeutic products, and applications. The Forum provides a venue to focus ongoing attention and visibility to these important drug development needs and facilitates exploration of new approaches across the drug development lifecycle. The Forum has held workshops that have contributed to the defining and establishment of regulatory science and have helped inform aspects of drug regulatory evaluation.

CLINICAL TRIALS AND CLINICAL PRODUCT DEVELOPMENT

Clinical research is the critical link between bench and bedside in developing new therapeutics. Significant infrastructural, cultural, and regulatory impediments challenge efforts to integrate clinical trials into the health care delivery system. Collaborative, cross-sector approaches can help articulate and address these key challenges and foster systemic responses. The Forum has convened a multiyear initiative to examine the state of clinical trials in the United States, identify areas of strength and weakness in our current clinical trial enterprise, and consider transformative strategies for enhancing the ways in which clinical trials are organized and conducted. In addition to sponsoring multiple symposia and workshops, under this initiative, the Forum is fostering innovative, collaborative efforts to facilitate needed change in areas such as improvement of clinical trial site performance.

INFRASTRUCTURE AND WORKFORCE FOR DRUG DISCOVERY, DEVELOPMENT, AND TRANSLATION

Considerable opportunities remain for enhancement and improvement of the infrastructure that supports the drug development enterprise. That infrastructure, which includes the organizational structure, framework, systems, and resources that facilitate the conduct of biomedical science for drug development, faces significant challenges. The science of drug discovery and development, and its translation into clinical practice, is cross-cutting and multidisciplinary. Career paths can be opaque or lack incentives such as recognition, career advancement, or financial security. The Forum has considered workforce needs as foundational to the advancement of drug discovery, development, and translation. It has convened workshops examining these issues, including consideration of strategies for developing a discipline of innovative regulatory science through the development of a robust workforce. The Forum will also host an initiative that will address needs for a workforce across the translational science lifecycle.

Project Staff

Carolyn Shore, Ph.D.
Forum Director
cshore@nas.edu
202-334-2669

Jennifer Hinners, M.D., M.P.H.
Program Officer

Amanda Wagner Gee, M.S.
Program Officer

Sylvia Ncha, M.P.H.
Associate Program Officer

Eeshan Khandekar, M.Sc.
Associate Program Officer

Melvin Joppy
Senior Program Assistant

For more information, please visit:

NATIONALACADEMIES.ORG/DRUGFORUM

Health and Medicine Division
Board on Health Sciences Policy

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The National Academies of SCIENCES ENGINEERING MEDICINE

Forum on NEUROSCIENCE and NERVOUS SYSTEM DISORDERS

The Forum on Neuroscience and Nervous System Disorders was established in 2006 to provide a venue for building partnerships, addressing challenges, and highlighting emerging issues related to brain disorders, which are common, major causes of premature mortality, and, in aggregate, the largest cause of disability worldwide. The Forum's meetings bring together leaders from government, industry, academia, disease advocacy organizations, and other interested parties to examine significant—and sometimes contentious—issues concerning scientific opportunities, priority setting, and policies related to research on neuroscience and brain disorders; the development, regulation, and use of interventions for the nervous system; and related ethical, legal, and social implications.

Forum members meet several times a year to exchange information, ideas, and differing perspectives. The Forum also sponsors workshops (symposia), workshop proceedings, and commissioned papers as additional mechanisms for informing its membership, other stakeholders, and the public about emerging issues and matters deserving scrutiny. Additional information is available at www.nas.edu/NeuroForum.

Recent Workshops

Biomarkers of Neuroinflammation (2017)

Enabling Novel Treatments for Nervous System Disorders by Improving Methods for Traversing the Blood-Brain Barrier (2017)

Accelerating Therapeutic Development for Pain and Opioid Use Disorders through Public-Private Partnerships (2017)

Neuroforensics: Exploring the Legal Implications of Emerging Neurotechnologies (2018)

Harnessing Digital Technology for Brain Disorders (2018)

Transgenic and Chimeric Neuroscience Research: Exploring the Scientific Opportunities Afforded by New Nonhuman Primate Models (2018)

The Role of Nonpharmacological Approaches to Pain Management (2018)

Advancing Gene-Targeted Therapies for Nervous System Disorders (2019)

Upcoming Workshop

Neuroscience Data in the Cloud (Sept 24, 2019)

Enhancing Scientific Reproducibility through Transparent Reporting (Sept 25-26, 2019)*

Challenges and a Way Forward in Sharing Clinical Trial Data (Nov 18-19, 2019)*

**Co-hosted with the Forum on Drug Discovery, Development, and Translation; National Cancer Policy Forum; and Roundtable on Genomics and Precision Health*

Staff

Clare Stroud, Forum Director
Sheena Posey Norris, Program Officer
Phoenix Wilson, Senior Program Assistant
Andrew Pope, Director, Board on Health Sciences Policy

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Roundtable on GENOMICS and PRECISION HEALTH

The sequencing of the human genome is rapidly opening new doors to research and progress in biology, medicine, and health care. At the same time, these developments have produced a diversity of new issues to be addressed.

The National Academies of Sciences, Engineering, and Medicine has convened a Roundtable on Genomics and Precision Health (previously the Roundtable on Translating Genomic-Based Research for Health) that brings together leaders from academia, industry, government, foundations and associations, and representatives of patient and consumer interests who have a mutual concern and interest in addressing the issues surrounding the translation of genome-based research for use in maintaining and improving health. The mission of the Roundtable is to advance the field of genomics and improve the translation of research findings to health care, education, and policy. The Roundtable will discuss the translation process, identify challenges at various points in the process, and discuss approaches to address those challenges.

The field of genomics and its translation involves many disciplines, and takes place within different economic, social, and cultural contexts, necessitating a need for increased communication and understanding across these fields. As a convening mechanism for interested parties from diverse perspectives to meet and discuss complex issues of mutual concern in a neutral setting, the Roundtable: fosters dialogue across sectors and institutions; illuminates issues, but does not necessarily resolve them; and fosters collaboration among stakeholders.

To achieve its objectives, the Roundtable conducts structured discussions, workshops, and symposia. Workshop summaries will be published and collaborative efforts among members are encouraged (e.g., journal articles). Specific issues and agenda topics are determined by the Roundtable membership, and span a broad range of issues relevant to the translation process.

Issues may include the integration and coordination of genomic information into health care and public health including encompassing standards for genetic screening and testing, improving information technology for use in clinical decision making, ensuring access while protecting privacy, and using genomic information to reduce health disparities. The patient and family perspective on the use of genomic information for translation includes social and behavioral issues for target populations. There are evolving requirements for the health professional community, and the need to be able to understand and responsibly apply genomics to medicine and public health.

Of increasing importance is the need to identify the economic implications of using genome-based research for health. Such issues include incentives, cost-effectiveness, and sustainability.

Issues related to the developing science base are also important in the translation process. Such issues could include studies of gene-environment interactions, as well as the implications of genomics for complex disorders such as addiction, mental illness, and chronic diseases.

Roundtable sponsors include federal agencies, pharmaceutical companies, medical and scientific associations, foundations, and patient/public representatives. For more information about the Roundtable on Genomics and Precision Health, please visit our website at nationalacademies.org/GenomicsRT or contact Sarah Beachy at 202-334-2217, or by e-mail at sbeachy@nas.edu.

500 Fifth Street, NW, Washington, DC 20001
Phone 202.334.2217 E-mail sbeachy@nas.edu nationalacademies.org/GenomicsRT

Roundtable on Genomics and Precision Health Membership

Geoffrey Ginsburg, M.D., Ph.D. (Co-Chair) Duke University
Michelle Penny, Ph.D. (Co-Chair) Biogen

Naomi Aronson, Ph.D.
BlueCross/BlueShield Association

Aris Baras, M.D., M.B.A.
Regeneron Pharmaceuticals

John Belmont, M.D., Ph.D.
Illumina

Karina Bienfait, Ph.D.
Merck and Co., Inc.

Vence Bonham, Jr., J.D.
National Human Genome Research Institute

Robert B. Darnell, M.D. Ph.D.
NY Genome Center / The Rockefeller University

Katherine Donigan, Ph.D.
U.S. Food and Drug Administration

W. Gregory Feero, M.D., Ph.D.
Journal of the American Medical Association

Jessica M. Gill, Ph.D., R.N., FAAN
National Institute of Nursing Research

Marc Grodman, M.D.
Genosity

Emily Harris, Ph.D., M.P.H.
National Cancer Institute

Richard Hodes, M.D.
National Institute on Aging

Praduman Jain, M.S.
Vibrent Health

Sekar Kathiresan, M.D.
Massachusetts General Hospital

Muin Khoury, M.D., Ph.D.
Centers for Disease Control and Prevention

Thomas Lehner, Ph.D., M.P.H.
National Institute of Mental Health

Patrick Loerch, Ph.D.
Johnson & Johnson

Sean McConnell, Ph.D.
American Medical Association

Mona Miller, M.P.P.
American Society of Human Genetics

Jennifer Moser, Ph.D.
U.S. Department of Veterans Affairs

Anna Pettersson, Ph.D.
Pfizer Inc.

Victoria M. Pratt, Ph.D., FACMG
Association for Molecular Pathology

Nadeem Sarwar, Ph.D.
Eisai Inc.

Sheri Schully, Ph.D.
NIH Office of Disease Prevention

Joan A. Scott, M.S., C.G.C.
Health Resources and Services Administration

Nikoletta Sidiropoulos, M.D.
University of Vermont Health Network Medical Group

Katherine Johansen Taber, Ph.D.
Myriad Women's Health

Jacquelyn Taylor, Ph.D.
New York University

Sharon Terry, M.A.
Genetic Alliance

Joyce Tung, Ph.D.
23andMe, Inc.

Jameson Voss, M.D.
U.S. Air Force

Michael S. Watson, Ph.D.
American College of Medical Genetics and Genomics

Karen Weck, M.D.
College of American Pathologists

Catherine A. Wicklund, M.S., C.G.C.
National Society of Genetic Counselors

Huntington F. Willard, Ph.D.
Geisinger Health

Janet K. Williams, Ph.D., R.N., FAAN
American Academy of Nursing

Sarah Wordsworth, Ph.D.
University of Oxford

Alicia Zhou, Ph.D.
Color Genomics

Project Staff
Sarah H. Beachy, Ph.D., *Roundtable Director*
Siobhan Addie, Ph.D., *Program Officer*
Meredith Hackmann, *Associate Program Officer*
Michael Berrios, M.A., *Senior Program Assistant*

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Planning Committee Biographies

CHAIR

HARVEY FINEBERG, M.D. PH.D., is president of the Gordon and Betty Moore Foundation. He previously served as president of the U.S. Institute of Medicine (now National Academy of Medicine), as provost of Harvard University, and as dean of the Harvard Chan School of Public Health. Prior to joining a philanthropic foundation, he devoted most of his academic career to the fields of health policy and medical decision-making. His past research has focused on global health, assessment of medical technology, evaluation and use of vaccines, and dissemination of medical innovations. Dr. Fineberg is a trustee of the Carnegie Endowment for International Peace and the China Medical Board. He previously served on the boards of the William and Flora Hewlett Foundation, the Josiah Macy, Jr. Foundation, and the Association François-Xavier Bagnoud (FXB) U.S.A. He is past chair of the boards of the Carnegie Endowment and the Hewlett Foundation. He helped found and served as president of the Society for Medical Decision Making. Dr. Fineberg serves on the editorial board of the *New England Journal of Medicine* and in a number of advisory capacities, including the foresight committee of the Veolia Environment Institute and the advisory board of the Peterson Center on Healthcare. Dr. Fineberg is co-author of the books *Clinical Decision Analysis*, *Innovators in Physician Education*, and *The Epidemic That Never Was*, an analysis of the controversial U.S. immunization program against swine flu in 1976. He has co-edited books on such diverse topics as AIDS prevention, vaccine safety, understanding risk in society, and global health and has authored numerous articles published in professional journals. Dr. Fineberg is the recipient of several honorary degrees, the Frank A. Calderone Prize in Public Health, the Henry G. Friesen International Prize in Health Research, and the Harvard Medal, awarded by the alumni association of the university from which he earned his bachelor's and doctoral degrees.

MEMBERS

OTIS BRAWLEY, M.D., M.A.C.P., F.A.S.C.O., F.A.C.E., is the Bloomberg Distinguished Professor of Oncology and Epidemiology at Johns Hopkins University. He is an authority on cancer screening and prevention and leads a broad research effort of cancer health disparities in the School of Medicine, the Bloomberg School of Public Health and the Sidney Kimmel Comprehensive Cancer Center, focusing on how to close racial, economic, and social disparities in the prevention, detection, and treatment of cancer. He is a member of the National Cancer Institute Board of Scientific Advisors and the National Academy of Medicine. Dr. Brawley is a graduate of University of Chicago, Pritzker School of Medicine. He completed an internal medicine residency at University Hospitals of Cleveland, Case-Western Reserve University, and a fellowship in medical oncology at the National Cancer Institute. He is board certified in internal medicine and medical oncology.

BARRY COLLER, M.D. serves as Physician–in–Chief, Vice President for Medical Affairs, and the David Rockefeller Professor at The Rockefeller University. An authority on the cardiovascular biology of integrins and transforming growth factor (TGF)-beta, Dr. Coller is a member of the Institute of Medicine (now National Academy of Medicine), the National Academy of Sciences, and was founding president of the Society for Clinical and Translational Science. Dr. Coller is a pioneer in the discovery and development of monoclonal antibodies for use as human therapeutics. He played the central role in discovering the active component and mechanism of abciximab (ReoPro(R)), and he was a leader in its subsequent clinical development resulting in one of the first FDA approvals of an antibody medicine. He has been a Member of Scientific Advisory Board at Scholar Rock, Inc. since September 2014.

STUART HOFFMAN, PH.D., is the point of contact for the Veterans Affairs (VA) Rehabilitation Research and Development Service (RR&D) program on traumatic brain injury (TBI). In this role, Dr. Hoffman has oversight for two VA TBI Research Centers and is the Co-chair of the Government Steering Committee for the VA/Department of Defense (DoD) Chronic Effects of Neurotrauma Consortium, as well as a the VA TBI subject matter expert for the National Research Action Plan for Improving Access to Mental Health Services for Veterans, Service Members, and Military Families. Dr. Hoffman also serves on several intra- and interagency advisory committees for VA and DoD, including the Congressionally-mandated Traumatic Brain Injury Advisory Committee for the Veterans Health Administration. Dr. Hoffman received his doctoral degree in behavioral and molecular neuroscience at Rutgers University in 1995 and completed his postdoctoral training in pharmacology at Virginia Commonwealth University in 1997. Dr. Hoffman's professional career began at Emory University as a Research Assistant Professor in the Department of Neurology in 1998 and was an Assistant Professor of Emergency Medicine from 2000 to 2006. Immediately prior to joining the VA in 2010, Dr. Hoffman was the Research Director for the Defense and Veterans Brain Injury Center in Johnstown, Pennsylvania. Dr. Hoffman has more than 30 years of translational neuroscience research experience that focus on neuroprotection and methods to promote recovery of function after brain injury.

VERONIQUE KIERMER, PH.D., is Executive Editor at PLOS, where she works closely with the editorial teams of the seven PLOS journals to continually improve the communication of research. Before joining PLOS in 2015, she was Executive Editor and Director of Author and Reviewer Services for Nature Publishing Group. In that capacity she oversaw editorial and research integrity policies across the Nature journals. She started her career in publishing as the founding Chief Editor of Nature Methods and subsequently took on publishing responsibility for the title and other online products. Dr. Kiermer has a Ph.D. in molecular biology from the Université Libre de Bruxelles, Belgium. Her postdoctoral work was in the laboratory of Dr. Eric Verdin at the Gladstone Institute, University of California, San Francisco, studying the transcriptional regulation of HIV. She worked on gene therapy projects at the biotechnology company Cell Genesys before moving into publishing in 2004. Since February 2017, she has served as Chair of the ORCID Board.

BENEDICT KOLBER, PH.D., serves as an Associate Professor in the Department of Biological Sciences and the Research and Education Coordinator for the Chronic Pain Research Consortium at Duquesne University. Dr. Kolber came to Duquesne in 2011 from Washington University in St. Louis. Dr. Kolber has over 15 years experience doing neuroscience research. His graduate work with Louis Muglia, M.D., Ph.D., looked at the role of the endocrine stress response in the modulation of stress adaptation, depression, and anxiety. His post-doctoral work with Robert Gereau IV, Ph.D., focused on understanding the role of the amygdala in the modulation of acute and chronic pain. Dr. Kolber also has extensive experience teaching at the undergraduate and graduate level. His teaching interests focus on teaching neuroscience, endocrinology, physiology, and in utilizing innovative teaching techniques in the classroom.

ALEXA MCCRAY, PH.D., is a Professor of Medicine at Harvard Medical School and the Department of Medicine, Beth Israel Deaconess Medical Center. She conducts research on knowledge representation and discovery, with a special focus on the significant problems that persist in the curation, dissemination, and exchange of scientific and clinical information in biomedicine and health. Dr. McCray is the former director of the Lister Hill National Center for Biomedical Communications, a research division of the National Library of Medicine at the National Institutes of Health. While at the NIH, she directed the design and development of a number of national information resources, including ClinicalTrials.gov. Before joining the NIH she was on the research staff of IBM's T.J. Watson Research Center. She received the PhD from Georgetown University, and for three years was on the faculty there. She conducted pre-doctoral research at the Massachusetts Institute of Technology. Dr. McCray joined Harvard Medical School in 2005, where she was founding co-director of the Center for Biomedical Informatics and associate director of the Francis A. Countway Library of Medicine. Dr. McCray was elected to the National Academy of Medicine in 2001. She is chair of the National Research Council's Board on Research Data and Information. She is a fellow of the American Association for the Advancement of Science, a fellow of the American College of Medical Informatics (ACMI), an honorary fellow of the International Medical Informatics Association, and a founding fellow of the International Academy of Health Sciences Informatics. She is a past president of

ACMI and a past member of the board of both the American Medical Informatics Association and the International Medical Informatics Association. She is a former editor-in-chief of Methods of Information in Medicine, and she is a past member of the editorial board of the Journal of the American Medical Informatics Association. She chaired the 2018 National Academies of Sciences, Engineering, and Medicine consensus study entitled Open Science by Design: Realizing a Vision for 21st Century Research.

JILL MESIROV, PH.D., is responsible for the overarching strategy for data science and research computing for health sciences at University of California (UC) San Diego School of Medicine. She is also a professor in the Department of Medicine. Dr. Mesirov is a computational scientist who has spent many years working in the area of high-performance computing on problems that arise in science, engineering, and business applications. Her research focuses on cancer genomics applying machine-learning methods to functional data derived from patient tumors. The lab analyzes this molecular data to determine the underlying biological mechanisms of specific tumor subtypes, to stratify patients according to their relative risks of relapse, and to identify candidate compounds for new treatments. In addition, Dr. Mesirov is committed to the development of practical, accessible software tools to bring these methods to the general biomedical research community. Her tools support almost 500,000 users worldwide. Before joining UC San Diego in 2015, Dr. Mesirov was associate director and chief informatics officer at the Broad Institute of MIT and Harvard, where she directed the Computational Biology and Bioinformatics Program. She previously served as manager of computational biology and bioinformatics in the Healthcare/Pharmaceutical Solutions Organization, director of research at Thinking Machines Corporation, and has also held positions in the mathematics department at the University of California, Berkeley and served as associate executive director of the American Mathematical Society. Dr. Mesirov received her B.A. in mathematics from the University of Pennsylvania and earned her M.A. and Ph.D. in mathematics from Brandeis University. She is a fellow of the American Association for the Advancement of Science (AAAS), the American Mathematical Society (AMS), the Association for Women in Mathematics, and the International Society for Computational Biology (ISCB).

MARTIN MURPHY, D.MED.SC, PH.D., F.A.S.C.O., is a member of the National Cancer Policy Forum of the National Academy of Medicine of the National Academy of Sciences; a Director of the Foundation for the National Institutes of Health. He is a Fellow of the American Society of Clinical Oncology (ASCO), and founding Executive Editor of the peer-reviewed biomedical journal: The Oncologist, Stem Cells and Stem Cells Translational Medicine. A co-founder of the Society for Translational Oncology; a member of the Scientific Advisory Board of Hatteras Venture Partners; a charter member of the International Advisory Board of the VU University Medical Imaging Center; a charter member of Queen's University Belfast School of Medicine International Review Board; Dr. Murphy is also Chairman Emeritus of the Conquer Cancer Foundation of ASCO; convener of ACT-China; and a steering committee member and senior consultant to the Chinese Society of Clinical Oncology. Dr. Murphy is founding Chief Executive Office of Project Data Sphere®, LLC, a non-profit enterprise devoted to cancer clinical trial data-transparency, data-sharing and data-analysis founded by the CEO Roundtable on Cancer's Life Sciences Consortium.

RICHARD NAKAMURA, PH.D. In 2013, Dr. Nakamura was named Director of the Center for Scientific Review (CSR) at the National Institutes of Health (NIH). He leads CSR's 450 scientists and administrative staff, overseeing their efforts to manage 80,000 incoming NIH grant applications a year and review the majority of them in CSR peer review groups. CSR holds about 1,500 review meetings a year, involving about 16,000 reviewers from the scientific community. Dr. Nakamura has had a 32-year tenure at the National Institute of Mental Health (NIMH), where he has served as both Scientific Director and Deputy Director of the institute, and he served as Acting Director from 2001 to 2002. During his time at NIMH, he received a number of leadership awards, including the Presidential Rank Award for outstanding leadership. He came to NIMH in 1976 as a postdoctoral fellow. In the mid-80's he coordinated NIMH's Biobehavioral Program and later was Chief of its Integrative Neuroscience Research Branch. Between 1997 and 2007, he served as the institute's Deputy Director. From 2007 to 2011 he was institute Scientific Director. While at NIMH, he also has held other positions, including Associate Director for Science Policy and Program Planning; Chief, Behavioral and Integrative Neuroscience Research Branch; and Coordinator, Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) Office of Animal Research Issues. Dr. Nakamura attended the Bronx High School of Science and earned his B.A. in psychology from Earlham College in Richmond, Indiana. He received his Ph.D. in psychology from the State University of New York in Stony Brook. Dr. Nakamura has expertise in a number of areas, including cognitive and comparative neuroscience, science policy/funding and ethics in science. He has published 30 peer reviewed scientific journal articles, most related to neurocognition in primates.

FRANKLIN SAYRE, M.L.I.S., currently serves as the STEM Librarian at Thompson Rivers University in British Columbia, Canada. Previously he served as a Health Sciences Librarian at the University of British Columbia and the University of Minnesota. Mr. Sayre has published and presented extensively on how academic librarians can support rigorous and reproducible research and has led initiatives supporting research commercialization and introductory computational research methods workshops for graduate students and faculty.

IDA SIM, PH.D., M.D., is a primary care physician, informatics researcher, and entrepreneur. She is a Professor of Medicine at the University of California, San Francisco, where she co-directs Biomedical Informatics at UCSF's Clinical and Translational Sciences Institute. Her current research focuses on the use of mobile apps and sensors to improve health and manage disease for populations and individuals, and to make clinical research faster and less expensive. She is a co-founder of Open mHealth, a non-profit organization that is breaking down barriers to mobile health app and data integration through an open software architecture. Open mHealth is an official working group of the Institute of Electrical and Electronics Engineers (IEEE) Standards Association and is on track to becoming a global IEEE standard. In 2005, Dr. Sim was the founding Project Coordinator of the World Health Organization's International Clinical Trials Registry Platform, where she led the establishment of the first global policy on clinical trial registration and defined the common 20-item Trial Registration Data Set that all registers worldwide adhere to. She served on the Institute of Medicine's committee on Strategies of Responsible Sharing of Clinical Trial Data and is a co-founder of Vivli, a non-profit organization that is building a global data sharing

platform for participant-level clinical trial data. Dr. Sim has served on multiple advisory committees on health information infrastructure for clinical care and research, including committees of the National Research Council and Institute of Medicine. She is a recipient of the United States Presidential Early Career Award for Scientists and Engineers (PECASE), a Fellow of the American College of Medical Informatics, and a member of the American Society for Clinical Investigation.

VALDA VINSON, PH.D., started her career in publishing when she joined the Science staff in 1999. Since then, she has handled research papers in the areas of structural biology, biochemistry, and biophysics as an Associate, Senior Editor, and Deputy Editor. She has also edited Perspectives and served as a team leader. As Editor, Dr. Vinson oversees research content in the areas of biological, life, and social sciences, working with these editors to attract and select exciting research papers and reviews, while maintaining high editorial standards. She earned an M.Sc. in Chemistry from Durban University in 1987 and a Ph.D. from Johns Hopkins University in 1992. Her postdoctoral studies were also undertaken at Johns Hopkins University, where she focused on structural and biochemical studies of cytoskeletal proteins. Before joining Science, she spent two years as a Senior Lecturer at the University of the Western Cape, South Africa.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Speaker Biographies

KEYNOTE SPEAKER

MARCIA MCNUTT, PH.D., is a geophysicist and president of the National Academy of Sciences. From 2013 to 2016, she served as editor-in-chief of the *Science* family of journals. Prior to joining *Science*, she was director of the U.S. Geological Survey (USGS) from 2009 to 2013. During her tenure, the USGS responded to a number of major disasters, including earthquakes in Haiti, Chile, and Japan, and the Deepwater Horizon oil spill. Dr. McNutt led a team of government scientists and engineers at BP headquarters in Houston who helped contain the oil and cap the well. She directed the flow rate technical group that estimated the rate of oil discharge during the spill's active phase. For her contributions, she was awarded the U.S. Coast Guard's Meritorious Service Medal.

Before joining the USGS, Dr. McNutt served as president and chief executive officer of the Monterey Bay Aquarium Research Institute (MBARI), in Moss Landing, California. During her time at MBARI, the institution became a leader in developing biological and chemical sensors for remote ocean deployment, installed the first deep-sea cabled observatory in U.S. waters, and advanced the integration of artificial intelligence into autonomous underwater vehicles for complex undersea missions. Dr. McNutt began her academic career at the Massachusetts Institute of Technology (MIT), where she was the E.A. Griswold Professor of Geophysics and directed the Joint Program in Oceanography/Applied Ocean Science & Engineering, jointly offered by MIT and the Woods Hole Oceanographic Institution. Her research area is the dynamics of the upper mantle and lithosphere on geologic time scales, work that has taken her to distant continents and oceans for field observations. She is a veteran of more than a dozen deep-sea expeditions, on most of which she was chief or co-chief scientist. Dr. McNutt received a B.A. in physics from Colorado College and her Ph.D. in Earth sciences at the Scripps Institution of Oceanography. Her honors include membership in the American Philosophical Society and the American Academy of Arts and Sciences. She holds honorary doctoral degrees from the Colorado College, the University of Minnesota, Monmouth University, and the Colorado School of Mines. In 1988, she was awarded the American Geophysical Union's Macelwane Medal for research accomplishments by a young scientist, and she received the Maurice Ewing Medal in 2007 for her contributions to deep-sea exploration. Dr. McNutt served as president of the American Geophysical Union (AGU) from 2000 to 2002. She was chair of the Board of Governors for Joint Oceanographic Institutions, responsible for operating the International Ocean Discovery Program's vessel JOIDES Resolution and associated research programs. She is a fellow of AGU, the Geological Society of America, American Association for the Advancement of Science, and International Association of Geodesy.

SPEAKERS

YARIMAR CARRASQUILLO, PH.D., joined the Pain and Integrative Neuroscience Branch of the National Center for Complementary and Integrative Health (NCCIH) at the National Institutes of Health (NIH) as an Investigator in 2014. In her lab, Dr. Carrasquillo directs a multifaceted, multidisciplinary research program focused on delineating the anatomical, molecular and cellular mechanisms that underlie pain perception and modulation. Dr. Carrasquillo received her B.S. in Biology from the University of Puerto Rico, Rio Piedras and her Ph.D. in Neuroscience from Baylor College of Medicine. She started her scientific career by studying the molecular basis of learning and memory as a Minority Biomedical Research Support Program (MBRS) Undergraduate Trainee in the lab of Dr. Sandra Peña de Ortiz. She continued studying the neural mechanisms underlying behavior during her graduate training in the lab of Dr. Robert W. Gereau at Baylor College of Medicine. Her graduate work revealed critical roles for the amygdala in the modulation of persistent pain and also demonstrated that the extracellular signal regulated kinase (ERK) plays a role in this process. Her postdoctoral studies in the lab of Dr. Jeanne Nerbonne at Washington University School of Medicine revealed previously unappreciated molecular and functional diversity of repolarizing voltage-gated K+ currents in central neurons.

ARTURO CASADEVALL M.D., PH.D., is a Bloomberg Distinguished Professor and Chair of the Molecular Microbiology and Immunology at Johns Hopkins School of Public Health. Previously, he served as Director, Division of Infectious Diseases at Montefiore Medical Center and Chair of the Department of Microbiology and Immunology at the Albert Einstein College of Medicine. He received his M.D. and Ph.D. degrees from New York University and completed his internship/residency in internal medicine at Bellevue Hospital. His infectious diseases training was at Montefiore and Einstein. The author of over 780 papers, numerous books and chapters, his major research interests are in fungal pathogenesis and the mechanisms of antibody action. He is also interested in the problems with scientific enterprise and with collaborators showed that misconduct accounts for the majority of retracted publications. He is editor-in-chief of mBio, Deputy Editor of the Journal of Clinical Investigation and serves on several numerous editorial boards. He has served on several NIH committees including the National Institute of Allergy and Infectious Diseases (NIAID) Strategic Plan, the Blue Ribbon Panel on Biodefense Research, the National Academy of Sciences panel that reviewed the FBI investigation on anthrax attacks, the NAS Federal Regulations and Reporting committee and the National Science Advisory Board for Biosecurity. He was a Commissioner in the National Commission on Forensic Science and previously served as President of the Medical Mycology Society of the Americas. He is currently the Chair the Board of Governors of the American Academy of Microbiology. He has received numerous honors including election to the American Society for Clinical Investigation, American Academy of Physicians, American Academy of Microbiology, Fellow of the American Academy for the Advancement of Science, American Academy of Arts and Sciences and the National Academy of Medicine.

AN-WEN CHAN, M.D., is a clinical epidemiologist and Mohs skin cancer surgeon at Women's College Hospital in Toronto, Canada. He is the Phelan Scientist at Women's College Research Institute, Associate Professor of Medicine at University of Toronto, and Director of Transplant Dermatology at University Health Network. After obtaining his doctorate as a Rhodes Scholar at University of Oxford, Dr. Chan served as Special Advisor to the Canadian Institutes of Health Research and helped coordinate the World Health Organization's International Clinical Trials Registry Platform. His research interests include the epidemiology and management of high-risk skin cancer in solid organ transplant recipients, as well as addressing issues of transparency and biases in clinical trials. He currently chairs the international Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) initiative to improve the quality of clinical trial protocols.

STEVEN GOODMAN, M.D., PH.D., is an Associate Dean of Clinical and Translational Research and Professor of Medicine and of Health Research & Policy, directing Stanford's Clinical and Translational Science Award (CTSA)/Spectrum training programs in medical research methods and serving as chief of the Division of Epidemiology in Health Research and Policy (HRP). He is co-founder and co-director of the Meta-research innovation Center at Stanford (METRICS), a group dedicated to examining and improving the reproducibility and efficiency of biomedical research. Dr. Goodman's own research concerns the proper measurement, conceptualization and synthesis of research evidence, with particular emphasis on Bayesian approaches to quantitation, and qualitative approaches arising from the philosophy of science. He is also interested in developing methods to use shared data to confirm and extend published science, as well as to explore new hypotheses. He also has worked on the connections between ethics and scientific methods, particularly in the domain of interventional research, and policy making. Finally, he has a strong interest in developing curricula and new models for teaching the foundations of good scientific practice, from question development to proper study design, conduct, analysis and inference. He has been a senior statistical editor of Annals of Internal Medicine since 1987 and was Editor of Clinical Trials: Journal of the Society for Clinical Trials from 2004-2013. He is Vice-chair of the Methodology Committee of the Patient Centered Outcomes Research Institute (PCORI), where he leads their open science and data sharing efforts, and is scientific advisor for the national Blue Cross–Blue Shield Technology Assessment Program. He has served on numerous Institute of Medicine committees since the mid 1990's, including chairing a 2012 committee on drug safety, and as a committee member on sharing data from clinical trials, whose report was released in January, 2015. Most recently, he served on an advisory group to the NIH director on the future of the National Library of Medicine, and was awarded the 2016 Spinoza Chair in Medicine from the University of Amsterdam. From 1989-2011, Steve served on the faculties of the Johns Hopkins Schools of Medicine and Public Health, where he was co-director of the doctoral program in Epidemiology and member (1989-2011) and then director (2007-2011) of the Johns Hopkins cancer center's Division of Biostatistics and Bioinformatics. At Hopkins, he taught courses on Systematic reviews and Meta-analysis, Diagnostic and prognostic testing, and several courses on epidemiologic, clinical research and inferential methods. He received an A.B. from Harvard University, majoring in Biochemistry and Applied Math, an M.D. from New York University, trained in pediatrics at Washington University in St. Louis, and received a master's degree in Biostatistics and Ph.D. in Epidemiology from Johns Hopkins University.

MAGALI HAAS, M.D., PH.D., founded Orion Bionetworks in July 2012 and serves as its Chief Executive Officer and President. Orion Bionetworks was transformed to Cohen Veterans Bioscience in 2015 to specifically dedicate research to improving the detection and treatment of post-traumatic stress (PTS) and traumatic brain injury (TBI) and related co-morbidities so that the burden of these conditions may be lessened on service members, veterans, and their families. Dr. Haas has over 15 years of pharmaceutical executive and clinical research experience, predominantly at Johnson & Johnson, where she assumed broad end-to-end development leadership roles in medical marketing, full clinical development, early development, and translational and biomarker sciences in psychiatry and neurology. She successfully filed new drug applications in the U.S. and Europe for risperidone indications in Autism, Adolescent Schizophrenia, Juvenile Bipolar Disorder, and Conduct Disorders. She also led Development Teams evaluating compounds for Depression, Neuropathic Pain, Epilepsy, and Migraine Disorder. She served 3 years as Chief Science and Technology Officer for One Mind for Research, a nonprofit organization launched in May 2012 by Patrick J. Kennedy. She orchestrated the launch of One Mind's seminal programs, Apollo, an informatics research portal, and Gemini, an international TBI/PTSD research program. As an "intrapreneur" at Johnson & Johnson, she established the first Neuroscience Translational Medicine & Integrative Solutions department, and co-founded the first Companion Diagnostics Center of Excellence as well as J&J's Healthcare Innovation team. She serves on several advisory boards including Brain Canada, Prophase, Pear Therapeutics, Partnership for Assessment and Accreditation of Scientific Practice (PAASP), and IMEC for nanoelectronics. Dr. Haas earned her B.S. in bioengineering from the University of Pennsylvania, an M.S. in biomedical engineering from Rutgers University, New Jersey, and her M.D., Ph.D. with distinction in neuroscience from Albert Einstein College of Medicine, New York.

MICHAEL KEISER, PH.D., is a Chan Zuckerberg Initiative Ben Barres Investigator and an Allen Frontiers Group Distinguished Investigator. He is an Assistant Professor at University of California San Francisco (UCSF) in the Dept. of Pharmaceutical Chemistry and the Institute for Neurodegenerative Diseases, with appointments in the Dept. of Bioengineering & Therapeutic Sciences and the Bakar Computational Health Sciences Institute. Before this, Dr. Keiser co-founded a startup bringing systems pharmacology methods he developed to the pharmaceutical industry and the U.S. Food and Drug Administration (FDA), where they are in use today. He holds multiple degrees from Stanford, including a B.Sc. in Computer Science. The Keiser lab combines machine learning and chemical biology methods to investigate how drug-like small molecules perturb protein networks to achieve their therapeutic effects.

MALCOLM MACLEOD, M.B. Ch.B., Ph.D., F.R.C.P. Ed. is a Professor in Neurology and Translational Neuroscience at the Centre for Clinical Brain Sciences, University of Edinburgh. After training in Internal Medicine, his Ph.D. concerned the neuroprotective actions of tacrolimus (FK506) and post-doctoral work in the Seckl lab defined a neuroprotective role for increased expression of the mineralocorticoid receptor. During a pivotal sabbatical year with Dr. Geoffrey Donnan at the National Stroke Research Institute in Melbourne, he began an involvement with stroke clinical trials, and it was in an effort to identify suitable drugs for such clinical trials that he began to develop techniques to allow the systematic review and meta-analysis of data from animal studies. In 2004 he founded, with Dr. David Howells, the Collaborative Approach to Meta-analysis and Review of Animal Data from Experimental Studies (CAMARADES, www.camarades.info). This approach has proved fruitful both in highlighting problems with in vivo

research and in providing evidence for clinical trial design. He was an author in *The Lancet* series on increasing value and reducing waste in biomedical research, and current work focusses on developing and evaluating strategies for research improvement. This includes evaluation of the impact of different strategies used by publishers to improve reporting. More recently he is a member of the recently formed U.K. Reproducibility Network; and Academic Coordinator of the European Quality In Preclinical Data (EQIPD) Innovative Medicines Initiative project. He chairs the Program Committee for the 2020 REWARD-EQUATOR conference hosted by QUEST Centre Berlin.

BRIAN NOSEK. PH.D., is co-Founder and Executive Director of the Center for Open Science (COS) (http://cos.io/) that operates the Open Science Framework (OSF) (http://osf.io/)--a collaborative management service for registering studies and archiving and sharing research materials and data. COS is enabling open and reproducible research practices worldwide. Dr. Nosek is also a Professor in the Department of Psychology at the University of Virginia. He received his Ph.D. from Yale University in 2002. He co-founded Project Implicit (http://projectimplicit.net/), a multi-university collaboration for research and education investigating implicit cognition--thoughts and feelings that occur outside of awareness or control. Dr. Nosek investigates the gap between values and practices, such as when behavior is influenced by factors other than one's intentions and goals. Research applications of this interest include implicit bias, decision-making, attitudes, ideology, morality, innovation, barriers to change, open science, and reproducibility. In 2015, he was named one of Nature's 10 and to the Chronicle for Higher Education Influence list.

MELISSA L. RETHLEFSEN, M.S.L.S., A.H.I.P., joined the University of Florida George A. Smathers Libraries as a tenured University Librarian, Associate Dean, and Fackler Director, Health Science Center Libraries in 2018. Prior to coming to the University of Florida, Ms. Rethlefsen was the Executive Director (Interim) of the Spencer S. Eccles Health Sciences Library at the University of Utah, where she also was the Section Director for the Center for Clinical and Translational Sciences' Systematic Review Core and the Director of the MidContinental Region of the National Network of Libraries of Medicine. Ms. Rethlefsen's recent work focuses on the reproducibility of systematic review search strategies. She is currently leading an international effort to create the Preferred Reporting Items for Systematic reviews and Meta-Analysis Search extension (PRISMA-S), a reporting guideline extension for search strategies. At the University of Utah, Ms. Rethlefsen spearheaded an institutional effort to enhance the reproducibility of research. This effort included two national conferences, Research Reproducibility 2016 and Research Reproducibility 2018: Building Research Integrity through Reproducibility. She also led the development of a year-long weekly interdisciplinary seminar series and formed a campus-wide Reproducibility Coalition to help influence change at the institutional level. She continues these efforts at the University of Florida, where she is planning to host the third Research Reproducibility conference, Educating for Reproducibility, in March 2020. She completed her undergraduate degree in History and English at the University of Minnesota, and her master's degree from the University of North Texas. She has also held positions at the Mayo Clinic Libraries, Minnesota Department of Health RN Barr Library, and the University of Minnesota Bio-Medical Library. She was a 2016/2017 Association of Academic Health Sciences Libraries/National Library of Medicine Leadership Fellow. For her work, Ms. Rethlefsen was named Estelle Brodman Academic Medical Librarian of the Year in 2015 by the Medical Library Association.

SHAI SILBERBERG, PH.D., is the Director of Research Quality at the NIH National Institute of Neurological Disorders and Stroke (NINDS) leading the Institute efforts to increase the excellence of science and the completeness of research reporting. In addition, Dr. Silberberg is a Program Director at NINDS overseeing basic research related to ion channels and transporters. Prior to joining NINDS, Dr. Silberberg was an Associate Professor at Ben-Gurion University of the Negev in Israel, investigating the biophysical functions and physiological roles of various ion channels

SOWMYA SWAMINATHAN, PH.D., is Head of Editorial Policy & Research Integrity for the Nature Research journals at Springer Nature where she is responsible for editorial policy development including in areas of research integrity and reproducibility. Prior to current position, Dr. Swaminathan was Chief Editor at Nature Cell Biology where she was responsible for developing content strategy, evolving and establishing editorial policy and peer review processes. She is committed to advancing transparency, quality, and integrity in scholarly publishing.

GEETA SWAMY, M.D., is an associate professor of Obstetrics & Gynecology, Division of Maternal-Fetal Medicine at Duke University and serves as associate vice provost and vice dean for scientific integrity for Duke University and the School of Medicine, respectively. Dr. Swamy works with leaders across campus to provide a consistent University vision for scientific integrity standards and expectations and drives efforts to ensure the advancement of a culture of integrity in research. She oversees the Duke Office of Scientific Integrity, which includes the Advancing Scientific Integrity, Services, & Training (ASIST) initiative, financial conflict of interest, clinical quality management program, misconduct in research, and the institutional research incident response committee. A highly accomplished clinician-scientist, Dr. Swamy specializes in perinatal infection and maternal immunization. She is the principal investigator on numerous grants from the NIH, CDC, and industry sponsors and has published more than 100 articles. She received her medical degree from the University of North Carolina at Chapel Hill and completed residency training in obstetrics and gynecology at the University of Pittsburgh and a fellowship in Maternal-Fetal Medicine at Duke University.

CARRIE WOLINETZ, PH.D., is Acting Chief of Staff and Associate Director for Science Policy at the Office of the Director (OD) at NIH. She advises the NIH Director on science policy matters of significance to the agency, the research community, and the public, on a wide range of issues including human subjects protections, biosecurity, biosafety, genomic data sharing, regenerative medicine, the organization and management of NIH, and the outputs and values of NIH-funded research. Prior to joining NIH, Dr. Wolinetz worked on biomedical research policy issues as the Deputy Director for Federal Affairs at the Association of American Universities (AAU) and the Director of Scientific Affairs and Public Relations at the Federation of American Societies for Experimental Biology (FASEB). She also served as the President of United for Medical Research, a leading NIH advocacy coalition. Outside of NIH, Dr. Wolinetz teaches as an Adjunct Assistant Professor at Georgetown University in the School of Foreign Service's program on Science, Technology & International Affairs. She has a B.S. in animal science from Cornell University, and she received her Ph.D. in animal science from The Pennsylvania State University, where her area of research was reproductive physiology.

DISCUSSION DOCUMENT

Selected Guidelines for Transparent Reporting

The purpose of this document is to help inform workshop discussions on improving the harmonization of guidelines for transparent reporting across journals and funding agencies so that biomedical researchers propose and report data in a consistent manner. This discussion document was compiled based on criteria described in the Transparency and Openness Promotion (TOP) guidelines (see https://cos.io/top/), the National Academies report on *Reproducibility and Replicability in Science* (see https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science), and the NIH policy on Enhancing Reproducibility through Rigor and Transparency (see https://grants.nih.gov/policy/reproducibility/index.htm). Each of the guidelines summarized below has a different scope and purpose.

Rather than a comprehensive comparison of the several guidelines, this document indicates the criteria related to transparent reporting that are covered by the various guidelines. It is intended as a background for the workshop discussion.

Brief Description of Guidelines Summarized

- The Animal Research: Reporting *In Vivo* Experiments (ARRIVE) guidelines recommend criteria for the reporting of primary research using animals. The guidelines were based on the CONSORT guidelines, but cover diverse study types. They were developed by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs); and were first published in PLOS Biology in 2010. The ARRIVE guidelines are currently being revised; a preprint of the revised ARRIVE guidelines and the accompanying Explanation and Elaboration document are available on BioRxiv, respectively at https://www.biorxiv.org/content/10.1101/703181v1 and at https://www.biorxiv.org/content/10.1101/703355v1.

- The Consolidated Standards of Reporting Trials (CONSORT) statement recommends information to include when reporting a randomized trial, was developed by an international group of trialists, methodologists, and medical journal editors; and was first published in JAMA in 1996 and last revised in 2010 and published in multiple journals. For more information, see http://www.consort-statement.org/.

- The ***DRAFT*** Materials, Design, Analysis, Reporting (MDAR) checklist for authors represents a generic set of minimum requirements applicable to all reporting studies in the life sciences for the explicit purpose of increasing transparent reporting and reproducibility, developed by the MDAR working group specifically seeking common reporting points from across multiple journals. The checklist is being pilot tested by volunteer journals, and therefore has not been published, at the time of this writing. The statement of task is available at https://osf.io/preprints/metaarxiv/9sm4x/.

- National Institutes of Health (NIH) policies:

    - "Data Sharing" policy, developed by the NIH, to encourage data generated with NIH funding be shared as widely and freely available as possible while still safeguarding participant privacy and confidential/proprietary data. The policy requires a data sharing plan for final

research data generated on grants of $500,000 or more, and was implemented in 2003. For more information, see https://grants.nih.gov/grants/policy/data_sharing/.

- "Dissemination of NIH-Funded Clinical Trial Information" policy, developed by the NIH and implemented in 2017, mandates that all clinical trials funded in part or in whole by NIH must be registered at ClinicalTrials.gov. The policy also requires that summary results be posted to ClinicalTrials.gov. For more information, see https://grants.nih.gov/grants/guide/notice-files/not-od-16-149.html.

- "Enhancing Reproducibility through Rigor and Transparency" policy, developed by the NIH, to clarify expectations for grantees and reviewers in describing or assessing proposed studies in applications and progress reports, announced in 2015 and implemented in 2016. For more information, see https://grants.nih.gov/policy/reproducibility/index.htm.

- The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement recommends an evidence-based minimum set of reporting elements for systematic reviews and meta-analyses; was developed by an international group, and was first published in 2009 in multiple journals (PLOS Medicine, Annals of Internal Medicine, BMJ, Journal of Clinical Epidemiology, and Open Medicine). For more information, see http://www.prisma-statement.org/.

- The Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement, developed by an international collaboration of trialists, methodologists, journal editors, and ethicists, recommends minimum content to include in clinical trial protocols, from study enrollment through closeout, first published in the Annals of Internal Medicine and BMJ in 2013. For more information, see: https://www.spirit-statement.org/

- The Transparency and Openness Promotion (TOP) guidelines describe eight modular standards of transparency that journals can select from to introduce policy and best practices at their publication; was developed by the Center for Open Science with input from journals, funders, and professional societies; and was first published in Science in 2015. See https://cos.io/top for information about TOP, including a summary table and links to complete policy language.

DISCUSSION DOCUMENT

Legend Criterion Covered Criterion Partially Covered Criterion Not Covered	ARRIVE *primary research in animals*	CONSORT *randomized trials*	***DRAFT*** MDAR *compilation of multiple life science journals*	NIH policies[1] *pre-clinical & clinical research*	PRISMA *meta-analyses*	SPIRIT *clinical trial protocols*	TOP Guidelines *modular publishing standards*
STUDY CHARACTERISTICS							
Code availability[2]							
Data availability[2,3]							
Data citation[2,3]							
STUDY METHODS							
Analytical methods: attrition, statistical precision, statistical power[2,3]							
Plan for analytical decisions/pre-registration[2,3]							
Animal use/sex as a biologic variable[4]							
Details of in-laboratory study replication[3]							
Details of study methods, computation, and associated parameters[2,3]							
Ethics[2, 5]							
Information on computational environment (e.g., operating system, library dependencies)[3]							
Materials availability discussed[2,3]							
Material authentication required[4]							
Methods and protocols[2,3,4]							
Sample definition[2]							

DISCUSSION DOCUMENT

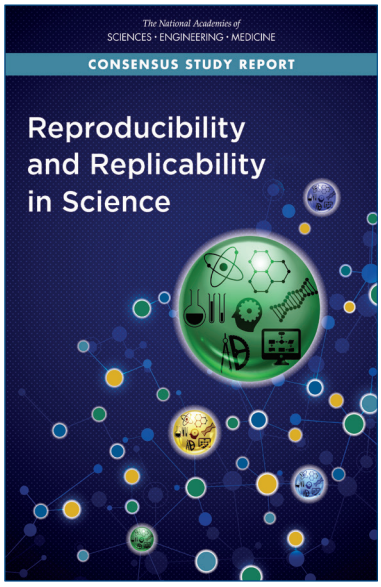
Legend Criterion Covered Criterion Partially Covered Criterion Not Covered	ARRIVE *primary research in animals*	CONSORT *randomized trials*	***DRAFT*** MDAR *compilation of multiple life science journals*	NIH policies[1] *pre-clinical & clinical research*	PRISMA *meta-analyses*	SPIRIT *clinical trial protocols*	TOP Guidelines *modular publishing standards*
RESULTS & DISCUSSION							
Adherence to community standards[2]							
Discussion of uncertainty[3]							
Discussion on generality constraints[3]							
Discuss/assess rigor of prior research[4]							
Dual Use Research of Concern*							

1. NIH policies represented include: Data Sharing Policy, Dissemination of NIH-Funded Clinical Trial Information, and Rigor and Transparency
2. Based on criteria described in the Transparency and Openness Promotion (TOP) guidelines, https://cos.io/top
3. Based on criteria described in the National Academies report, *Reproducibility and Replicability in Science*
4. Based on criteria described in the NIH policy on Enhancing Reproducibility through Rigor and Transparency
5. Covered by multiple NIH policies and offices, including Human Subjects Research policies and the NIH Office of Laboratory Animal Welfare

* Covered by US Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern

Consensus Study Report
May 2019
HIGHLIGHTS

Reproducibility and Replicability in Science



One of the pathways by which the scientific community confirms the validity of a new scientific discovery is by repeating the research that produced it. When a scientific effort fails to independently confirm the computations or results of a previous study, some fear that it may be a symptom of a lack of rigor in science, while others argue that such an observed inconsistency can be an important precursor to new discovery.

Concerns about reproducibility and replicability have been expressed in both scientific and popular media. As these concerns came to light, Congress requested that the National Academies of Sciences, Engineering, and Medicine conduct a study to assess the extent of issues related to reproducibility and replicability and to offer recommendations for improving rigor and transparency in scientific research.

The National Academies' study resulted in a report, *Reproducibility and Replicability in Science* (2019), that offers definitions of reproducibility and replicability and examines the factors that may lead to non-reproducibility and non-replicability in research. While reproducibility is straightforward and should generally be expected, the report says, replicability is more nuanced, and in some cases a lack of replicability can aid the process of scientific discovery. The report provides recommendations to researchers, academic institutions, journals, and funders on steps they can take to improve reproducibility and replicability in science.

DEFINING REPRODUCIBILITY AND REPLICABILITY

The terms "reproducibility" and "replicability" are often used interchangeably, but the report uses each term to refer to a separate concept.

Reproducibility means computational reproducibility—obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis.

Replicability means obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

In short, reproducing research involves using the *original* data and code, while replicating research involves *new* data collection and similar methods used by previous studies. These two processes also differ in the type of results that should be expected. In general, when a researcher transparently reports a study and makes available the underlying digital artifacts, such as data and code, the results should be computationally reproducible. In contrast, even when a study was rigorously conducted according to best practices, correctly analyzed, and transparently reported, it may fail to be replicated.

REPRODUCIBILITY

The committee's definition of reproducibility focuses on computation because most scientific and engineering research disciplines use computation as a tool, and the abundance of data and widespread use of computation have transformed many disciplines. However, this revolution is not yet uniformly reflected in how scientists use software and how scientific results are published and shared. These shortfalls have implications for reproducibility, because scientists who wish to reproduce research may lack the information or training they need to do so.

When results are produced by complex computational processes using large volumes of data, the traditional methods section of a scientific paper is insufficient to convey the necessary information for others to reproduce the results. Additional information related to data, code, models, and computational analysis is needed.

To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis.

If sufficient data, code, and description of methods are available and a second researcher follows the methods described by the first researcher, one expects in many cases full bitwise reproduction of the original results—obtaining the same exact numeric values. For some research questions, bitwise reproduction may not be attainable and reproducible results could be obtained within an accepted range of variation.

How common is non-reproducibility in research? The evidence base is incomplete, and determining the extent of issues related to computational reproducibility across or within fields of science would be a massive undertaking with a low probability of success. However, a number of systematic efforts to reproduce computational results across a variety of fields have failed in more than half of the attempts made—mainly due to insufficient detail on digital artifacts such as data, code, and computational workflow.

REPLICABILITY

One important way to confirm or build on previous results is to follow the same methods, obtain new data, and see if the results are consistent with the original. A successful replication does not guarantee that the original scientific results of a study were correct, however, nor does a single failed replication conclusively refute the original claims.

Non-replicability can arise from a number of sources. The committee classified sources of non-replicability into those that are potentially helpful to gaining knowledge, and those that are unhelpful.

Potentially helpful sources of non-replicability include inherent but uncharacterized uncertainties in the system being studied. These sources of non-replicability are a normal part of the scientific process, due to the intrinsic variation or complexity in nature, the scope of current scientific knowledge, and the limits of our current technologies. In such cases, a failure to replicate may lead to the discovery of new phenomena or new insights about variability in the system being studied.

In other cases, non-replicability is due to shortcomings in the design, conduct, and communication of a study. Whether arising from lack of knowledge, perverse incentives, sloppiness, or bias, these unhelpful sources of non-replicability reduce the efficiency of scientific progress.

Unhelpful sources of non-replicability can be minimized through initiatives and practices aimed at improving research design and methodology through training and mentoring, repeating experiments before publication,

rigorous peer review, utilizing tools for checking analysis and results, and better transparency in reporting. Efforts to minimize avoidable and unhelpful sources of non-replicability warrant continued attention.

Researchers who knowingly use questionable research practices with the intent to deceive are committing misconduct or fraud. It can be difficult in practice to differentiate between honest mistakes and deliberate misconduct, because the underlying action may be the same while the intent is not. Scientific misconduct in the form of misrepresentation and fraud is a continuing concern for all of science, even though it accounts for a very small percentage of published scientific papers.

IMPROVING REPRODUCIBILITY AND REPLICABILITY IN RESEARCH

The report recommends a range of steps that stakeholders in the research enterprise should take to improve reproducibility and replicability, including:

- All researchers should include a clear, specific, and complete description of how the reported results were reached. Reports should include details appropriate for the type of research, including:
  - ◊ a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study;
  - ◊ a clear description of the analysis of data and decisions for exclusion of some data or inclusion of other;
  - ◊ for results that depend on statistical inference, a description of the analytic decisions and when these decisions were made and whether the study is exploratory or confirmatory;
  - ◊ a discussion of the expected constraints on generality, such as which methodological features the authors think could be varied without affecting the result and which must remain constant;
  - ◊ reporting of precision or statistical power; and
  - ◊ discussion of the uncertainty of the measurements, results, and inferences.
- Funding agencies and organizations should consider investing in research and development of open-source, usable tools and infrastructure that support reproducibility for a broad range of studies across different domains in a seamless fashion. Concurrently, investments would be helpful in outreach to inform and train researchers on best practices and how to use these tools.
- Journals should consider ways to ensure computational reproducibility for publications that make claims based on computations, to the extent ethically and legally possible.
- The National Science Foundation should take steps to facilitate the transparent sharing and availability of digital artifacts, such as data and code, for NSF-funded studies—including developing a set of criteria for trusted open repositories to be used by the scientific community for objects of the scholarly record, and endorsing or considering the creation of code and data repositories for long-term archiving and preservation of digital artifacts that support claims made in the scholarly record based on NSF-funded research, among other actions.

Additional recommendations, along with detail on those included above, can be found in the report.

CONFIDENCE IN SCIENCE

Replicability and reproducibility, useful as they are in building confidence in scientific knowledge, are not the only ways to gain confidence in scientific results. Multiple channels of evidence from a variety of studies provide a robust means for gaining confidence in scientific knowledge over time. Research synthesis and meta-analysis, for example, are valuable methods for assessing the reliability and validity of bodies of research. A goal of science is to understand the overall effect from a set of scientific studies, not to strictly determine whether any one study has replicated any other.

The committee was asked to consider if lack of replication and reproducibility impacts the public's perception of science. The committee was not aware of data that would directly answer that question, and coverage of the issue in public media remains low. Regardless, the report notes that scientists and journalists bear responsibility

for misrepresentation in the public's eye when they overstate or otherwise misrepresent the implications of scientific research. The report offers the following recommendations:

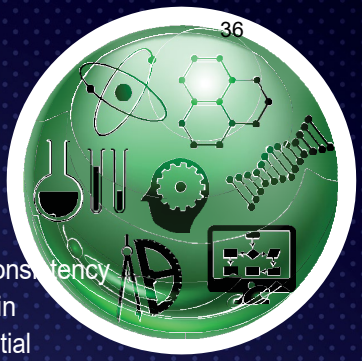
- Scientists should take care to avoid overstating the implications of their research, exercised also in their review of press releases, especially when the results bear directly on matters of keen public interest and possible action.
- Journalists should report on scientific results with as much context and nuance as the medium allows. In covering issues related to replicability and reproducibility, journalists should help their audiences understand the differences between non-reproducibility and non-replicability due to fraudulent conduct of science, and instances in which the failure to reproduce or replicate may be due to evolving best practices in methods or inherent uncertainty in science.

COMMITTEE ON REPRODUCIBILITY AND REPLICABILITY IN SCIENCE

HARVEY V. FINEBERG (NAM), (*Chair*), Gordon and Betty Moore Foundation; **DAVID B. ALLISON** (NAM), School of Public Health-Bloomington Director, Indiana University; **LORENA A. BARBA,** School of Engineering and Applied Science, George Washington University; **DIANNE CHONG** (NAE), Boeing Research and Technology (retired); **JULIANA FREIRE,** Tandon School of Engineering, New York University; **GERALD GABRIELSE** (NAS), Department of Physics, Northwestern University; **CONSTANTINE GATSONIS,** Center for Statistical Sciences, Brown University; **EDWARD HALL,** Department of Philosophy, Harvard University; **THOMAS H. JORDAN** (NAS), Department of Earth Sciences, University of Southern California; **DIETRAM A. SCHEUFELE,** Madison and Morgridge Institute for Research, University of Wisconsin–Madison; **VICTORIA STODDEN,** Institute for Data Sciences and Engineering, University of Illinois at Urbana–Champaign; **TIMOTHY D. WILSON,** Department of Psychology, University of Virginia; **WENDY WOOD,** Department of Psychology, University of Southern California and INSEAD-Sorbonne University; **JENNIFER HEIMBERG,** *Study Director;* **THOMAS ARRISON**, *Program Director;* **MICHAEL COHEN**, *Senior Program Officer;* **MICHELLE SCHWALBE**, *Director;* **TINA WINTERS**, *Associate Program Officer;* **THELMA COX,** *Program Coordinator*.

For More Information . . . This Consensus Study Report Highlights was prepared by the Board on Behavioral, Cognitive, and Sensory Sciences based on the Consensus Study Report *Reproducibility and Replicability in Science* (2019). The study was sponsored by the Alfred P. Sloan Foundation and the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project. Copies of the Consensus Study Report are available from the National Academies Press, (800) 624-6242; http://www.nas.edu/ReproducibilityinScience.

10 Things to Know About
Reproducibility and Replicability



One of the pathways by which the scientific community confirms the validity of a new scientific discovery is by repeating the research that produced it. When a scientific effort fails to independently confirm the computations or results of a previous study, some argue that such an observed inconsistency can be an important precursor to new discovery while others fear it may be a symptom of a lack of rigor in science. When a newly reported scientific study has far-reaching implications for science or a major, potential impact on the public, the question of its reliability takes on heightened importance.

1. **The terms reproducibility and replicability take on a range of meanings in contemporary usage.** The report distinguishes and defines the terms as follows: **Reproducibility** means obtaining consistent results using the same input data, computational steps, methods, and conditions of analysis; it is synonymous with computational reproducibility. **Replicability** means obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

2. **Reproducibility and replicability matter.** Reproducibility and replicability are often cited as hallmarks of good science. Being able to reproduce the computational results of another researcher starting with the same data and replicating a previous study to test its results facilitate the self-correcting nature of science.

3. **Computational reproducibility is more prominent now than ever because of the growth in reliance on computing across all of science.** When a researcher reports a study and makes the underlying data and code available, those results should be computationally reproducible by another researcher.

4. **A successful replication does not guarantee that the original scientific results of a study were correct, nor does a single failed replication conclusively refute the original claims.** Unlike the typical expectation of reproducibility between two computations, expectations about replicability are more nuanced.

5. **Occasionally, non-replicability may be caused by helpful sources that advance scientific knowledge, such as discovering previously unknown effects or sources of variability. At other times, a study cannot be replicated due to unhelpful sources, ranging from simple mistakes to methodological errors to bias and fraud.**

6. **Not all studies can be replicated.** While scientists are able to test for replicability of most studies, it is impossible to do so for studies of ephemeral phenomena.

7. **One type of scientific research tool, statistical inference, has an outsized role in replicability discussions due to the frequent misuse of statistics and the use of a *p*-value threshold for determining "statistical significance."** Biases in published research can occur due to the excess reliance on and misunderstanding of statistical significance.

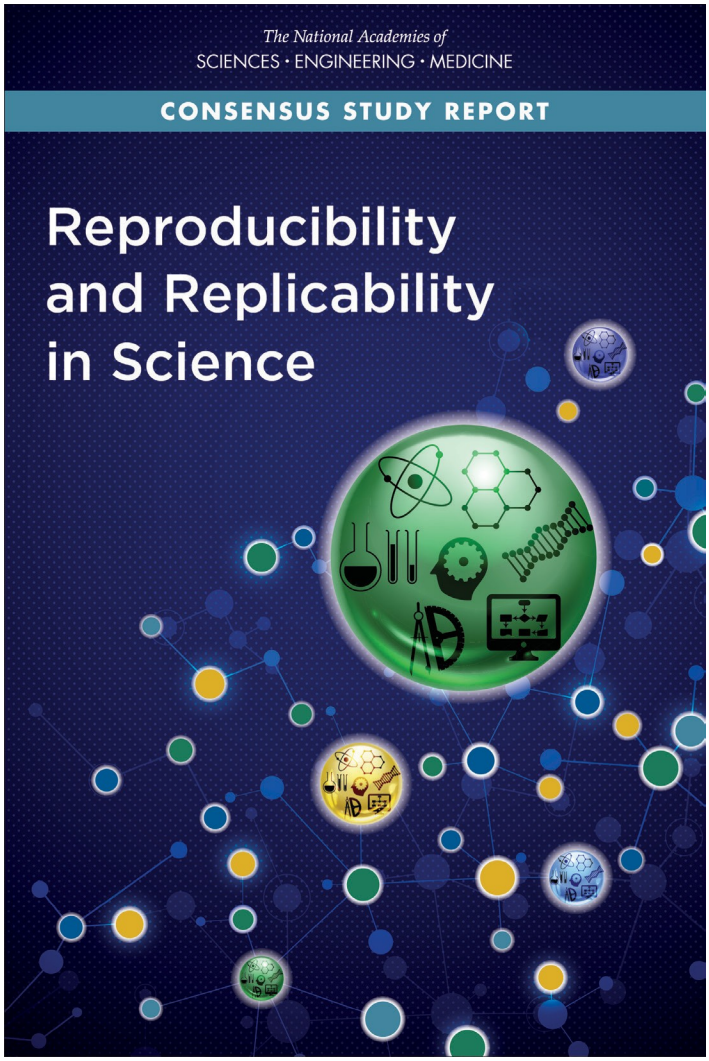
8. **Examining replicability becomes especially important when new findings have strong implications for individual health and well-being, policy choices, or the future course of scientific research.**

9. **Beyond reproducibility and replicability, systematic reviews and syntheses of scientific evidence are among the important ways to gain confidence in scientific results.**

10. **Academic institutions, journals, conference organizers, funding organizations, and policy makers can all play a role in improving the reproducibility and replicability of research. Responsibility begins with researchers, who should take care to estimate and explain the uncertainty inherent in their results and inferences, make proper use of statistical methods, and describe their methods and data in a clear, accurate, and complete way.**

Reproducibility and Replicability in Science **is available at www.nap.edu/25303.**

The National Academies of SCIENCES ENGINEERING MEDICINE



http://nationalacademies.org/reproducibilityinscience

Committee Members

HARVEY V. FINEBERG, (*Chair*), Gordon and Betty Moore Foundation

DAVID B. ALLISON, School of Public Health-Bloomington, Indiana University

LORENA A. BARBA, School of Engineering and Applied Science, George Washington University

DIANNE CHONG, Boeing Research and Technology (Retired)

JULIANA FREIRE, Tandon School of Engineering, New York University

GERALD GABRIELSE, Department of Physics, Northwestern University

CONSTANTINE GATSONIS, Center for Statistical Sciences, Brown University

EDWARD HALL, Department of Philosophy, Harvard University

THOMAS H. JORDAN, Department of Earth Sciences, University of Southern California

DIETRAM A. SCHEUFELE, Madison and Morgridge Institute for Research, University of Wisconsin-Madison

VICTORIA STODDEN, Institute for Data Sciences and Engineering, University of Illinois at Urbana-Champaign

TIMOTHY D. WILSON, Department of Psychology, University of Virginia

WENDY WOOD, Department of Psychology, University of Southern California and INSEAD-Sorbonne University

The study is sponsored by the National Science Foundation and The Alfred P. Sloan Foundation.

OPEN ACCESS Freely available online

Perspective

Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research

Carol Kilkenny[1]*, William J. Browne[2], Innes C. Cuthill[3], Michael Emerson[4], Douglas G. Altman[5]

1 The National Centre for the Replacement, Refinement and Reduction of Animals in Research, London, United Kingdom, **2** School of Veterinary Science, University of Bristol, Bristol, United Kingdom, **3** School of Biological Sciences, University of Bristol, Bristol, United Kingdom, **4** National Heart and Lung Institute, Imperial College London, United Kingdom, **5** Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom

In the last decade the number of bioscience journals has increased enormously, with many filling specialised niches reflecting new disciplines and technologies. The emergence of open-access journals has revolutionised the publication process, maximising the availability of research data. Nevertheless, a wealth of evidence shows that across many areas, the reporting of biomedical research is often inadequate, leading to the view that even if the science is sound, in many cases the publications themselves are not "fit for purpose," meaning that incomplete reporting of relevant information effectively renders many publications of limited value as instruments to inform policy or clinical and scientific practice [1–21]. A recent review of clinical research showed that there is considerable cumulative waste of financial resources at all stages of the research process, including as a result of publications that are unusable due to poor reporting [22]. It is unlikely that this issue is confined to clinical research [2–14,16–20].

Failure to describe research methods and to report results appropriately therefore has potential scientific, ethical, and economic implications for the entire research process and the reputation of those involved in it. This is particularly true for animal research, one of the most controversial areas of science. The largest and most comprehensive review of published animal research undertaken to date, to our knowledge, has highlighted serious omissions in the way research using animals is reported [5]. The survey, commissioned by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), a UK Government-sponsored scientific organisation, found that only 59% of the 271 randomly chosen articles assessed stated the hypothesis or objective of the study, and the number and characteristics of the animals used (i.e., species/strain, sex, and age/weight). Most of the papers surveyed did not report using randomisation (87%) or blinding (86%) to reduce bias in animal selection and outcome assessment. Only 70% of the publications that used statistical methods fully described them and presented the results with a measure of precision or variability [5]. These findings are a cause for concern and are consistent with reviews of many research areas, including clinical studies, published in recent years [2–22].

Good Reporting Is Essential for Peer Review and to Inform Future Research

Scrutiny by scientific peers has long been the mainstay of "quality control" for the publication process. The way that experiments are reported, in terms of the level of detail of methods and the presentation of key results, is crucial to the peer review process and, indeed, the subsequent utility and validity of the knowledge base that is used to inform future research. The onus is therefore on the research community to ensure that their research articles include all relevant information to allow in-depth critique, and to avoiding duplicating studies and performing redundant experiments. Ideally scientific publications should present sufficient information to allow a knowledgeable reader to understand what was done, why, and how, and to assess the biological relevance of the study and the reliability and validity of the findings. There should also be enough information to allow the experiment to be repeated [23]. The problem therefore is how to ensure that all relevant information is included in research publications.

Using Reporting Guidelines Measurably Improves the Quality of Reporting

Evidence provided by reviews of published research suggests that many researchers and peer reviewers would benefit from guidance about what information should be provided in a research article. The CONSORT Statement for randomised controlled clinical trials was one of the first guidelines developed in response to this need [24,25]. Since publication, an increasing number of leading journals have supported CONSORT as part of their instructions to authors [26,27]. As a result, convincing evidence is emerging that CONSORT improves the quality and transparency of reports of clinical trials [28,29].

Following CONSORT, many other guidelines have been developed—there are currently more than 90 available for reporting different types of health research, most of which have been published in the last ten years (see http://www.equator-network.org and references [30,31]). Guidelines have also been developed to improve the reporting of other specific bioscience research areas includ-

The Perspective section provides experts with a forum to comment on topical or controversial issues of broad interest.

Citation: Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010) Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biol 8(6): e1000412. doi:10.1371/journal.pbio.1000412

Published June 29, 2010

Funding: This project was initiated, funded, and led by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs).

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: ARRIVE, Animals in Research: Reporting *In Vivo* Experiments; NC3Rs, National Centre for the Replacement, Refinement and Reduction of Animals in Research

* E-mail: carol.kilkenny@nc3rs.org.uk

ing metabolomics and gene expression studies [32–37]. Several organisations support the case for improved reporting and recommend the use of reporting guidelines, including the International Committee of Medical Journal Editors, the Council of Science Editors, the Committee on Publication Ethics, and the Nuffield Council for Bioethics [38–41].

Improving the Reporting of Animal Experiments—The ARRIVE Guidelines

Most bioscience journals currently provide little or no guidance on what information to report when describing animal research [42–50]. Our review found that 4% of the 271 journal articles assessed did not report the number of animals used anywhere in the methods or the results sections [5]. Reporting animal numbers is essential so that the biological and statistical significance of the experimental results can be assessed or the data reanalysed, and is also necessary if the experimental methods are to be repeated. Improved reporting of these and other details will maximise the availability and utility of the information gained from every animal and every experiment, preventing unnecessary animal use in the future. To address this, we led an initiative to produce guidelines for reporting animal research. The guidelines, referred to as ARRIVE (Animals in Research: Reporting *In Vivo* Experiments), have been developed using the CONSORT Statement as their foundation [24,25].

The ARRIVE guidelines consist of a checklist of 20 items describing the minimum information that all scientific publications reporting research using animals should include, such as the number and specific characteristics of animals used (including species, strain, sex, and genetic background); details of housing and husbandry; and the experimental, statistical, and analytical methods (including details of methods used to reduce bias such as randomisation and blinding). All the items in the checklist have been included to promote high-quality, comprehensive reporting to allow an accurate critical review of what was done and what was found.

Consensus and consultation are the corner-stones of the guideline development process [51]. To maximise their utility, the ARRIVE guidelines have been prepared in consultation with scientists, statisticians, journal editors, and research funders. We convened an expert working group, comprising researchers and statisticians from a range of disciplines, and journal editors from *Nature Cell Biology*, *Science*, *Laboratory Animals*, and the *British Journal of Pharmacology* (see Acknowledgments). At a one-day meeting in June 2009, the working group agreed the scope and broad content of a draft set of guidelines that were then used as the basis for a wider consultation with the scientific community, involving researchers, and grant holders and representatives of the major bioscience funding bodies including the Medical Research Council, Wellcome Trust, Biotechnology and Biological Sciences Research Council, and The Royal Society (see Table 1). Feedback on the content and wording of the items was incorporated into the final version of the checklist. Further feedback on the content utility of the guidelines is encouraged and sought.

The ARRIVE guidelines (see Table 2) can be applied to any area of bioscience research using laboratory animals, and the inherent principles apply not only to reporting comparative experiments but also to other study designs. Laboratory animal refers to any species of animal undergoing an experimental procedure in a research laboratory or formal test setting. The guidelines are not intended to be mandatory or absolutely prescriptive, nor to standardise or formalise the structure of reporting. Rather they provide a checklist that can be used to guide authors preparing manuscripts for publication, and by those involved in peer review for quality assurance, to ensure completeness and transparency.

Improved Reporting Will Maximise the Output of Published Research

These guidelines were developed to maximise the output from research using animals by optimising the information that is provided in publications on the design, conduct, and analysis of the experiments.

Table 1. Funding bodies consulted.

Name of Bioscience Research Funding Body
Medical Research Council
Biotechnology and Biological Sciences Research Council
Wellcome Trust
The Royal Society
Association of Medical Research Charities
British Heart Foundation
Parkinson's Disease Society

doi:10.1371/journal.pbio.1000412.t001

The need for such guidelines is further illustrated by the systematic reviews of animal research that have been carried out to assess the efficacy of various drugs and interventions in animal models [8,9,13,52–55]. Well-designed and -reported animal studies are the essential building blocks from which such a systematic review is constructed. The reviews have found that, in many cases, reporting omissions, in addition to the limitations of the animal models used in the individual studies assessed in the review, are a barrier to reaching any useful conclusion about the efficacy of the drugs and interventions being compared [2,3].

Driving improvements in reporting research using animals will require the collective efforts of authors, journal editors, peer reviewers, and funding bodies. There is no single simple or rapid solution, but the ARRIVE guidelines provide a practical resource to aid these improvements. The guidelines will be published in several leading bioscience research journals simultaneously [56–60], and publishers have already endorsed the guidelines by including them in their journal Instructions to Authors subsequent to publication. The NC3Rs will continue to work with journal editors to extend the range of journals adopting the guidelines, and with the scientific community to disseminate the guidelines as widely as possible (http://www.nc3rs.org.uk/ARRIVE).

Acknowledgments

The NC3Rs gratefully acknowledges the expertise and advice that all the contributors have given to developing the guidelines. We would particularly like to acknowledge the contribution of the other members of NC3Rs Reporting Guidelines Working Group (note that the working group members and authors who contributed to these guidelines were advising in their personal capacity and their input does not necessarily represent the policy of the organisations with which they are associated): Professor David Balding, Department of Epidemiology & Public Health, Imperial College, London UK; Dr Colin Dunn Editor *Laboratory Animals* (Royal Society of Medicine press); Dr. Stella Hurtley, Senior Editor *Science*; Professor Ian McGrath Editor-in-Chief *British Journal of Pharmacology* (Wiley Blackwell publishers); and Dr. Clare Stanford, Department of Psychopharmacology, University College, London UK. We would also like to thank NC3Rs grant holders, the Medical Research Council, Biotechnology and Biological Sciences Research Council, Wellcome Trust, Parkinson's Disease Society, British Heart Foundation and their grant holders and funding committee members who provided feedback on the guidelines; and Dr. Kathryn Chapman and Dr. Vicky Robinson (both NC3Rs) for their help with the manuscript.

Table 2. Animal Research: Reporting *In Vivo* experiments: The ARRIVE guidelines.

	ITEM	RECOMMENDATION
TITLE	1	Provide as accurate and concise a description of the content of the article as possible.
ABSTRACT	2	Provide an accurate summary of the background, research objectives (including details of the species or strain of animal used), key methods, principal findings, and conclusions of the study.
INTRODUCTION		
Background	3	a. Include sufficient scientific background (including relevant references to previous work) to understand the motivation and context for the study, and explain the experimental approach and rationale. b. Explain how and why the animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology.
Objectives	4	Clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested.
METHODS		
Ethical statement	5	Indicate the nature of the ethical review permissions, relevant licences (e.g. Animal [Scientific Procedures] Act 1986), and national or institutional guidelines for the care and use of animals, that cover the research.
Study design	6	For each experiment, give brief details of the study design, including: a. The number of experimental and control groups. b. Any steps taken to minimise the effects of subjective bias when allocating animals to treatment (e.g., randomisation procedure) and when assessing results (e.g., if done, describe who was blinded and when). c. The experimental unit (e.g. a single animal, group, or cage of animals). A time-line diagram or flow chart can be useful to illustrate how complex study designs were carried out.
Experimental procedures	7	For each experiment and each experimental group, including controls, provide precise details of all procedures carried out. For example: a. How (e.g., drug formulation and dose, site and route of administration, anaesthesia and analgesia used [including monitoring], surgical procedure, method of euthanasia). Provide details of any specialist equipment used, including supplier(s). b. When (e.g., time of day). c. Where (e.g., home cage, laboratory, water maze). d. Why (e.g., rationale for choice of specific anaesthetic, route of administration, drug dose used).
Experimental animals	8	a. Provide details of the animals used, including species, strain, sex, developmental stage (e.g., mean or median age plus age range), and weight (e.g., mean or median weight plus weight range). b. Provide further relevant information such as the source of animals, international strain nomenclature, genetic modification status (e.g. knock-out or transgenic), genotype, health/immune status, drug- or test-naïve, previous procedures, etc.
Housing and husbandry	9	Provide details of: a. Housing (e.g., type of facility, e.g., specific pathogen free (SPF); type of cage or housing; bedding material; number of cage companions; tank shape and material etc. for fish). b. Husbandry conditions (e.g., breeding programme, light/dark cycle, temperature, quality of water etc. for fish, type of food, access to food and water, environmental enrichment). c. Welfare-related assessments and interventions that were carried out before, during, or after the experiment.
Sample size	10	a. Specify the total number of animals used in each experiment and the number of animals in each experimental group. b. Explain how the number of animals was decided. Provide details of any sample size calculation used. c. Indicate the number of independent replications of each experiment, if relevant.
Allocating animals to experimental groups	11	a. Give full details of how animals were allocated to experimental groups, including randomisation or matching if done. b. Describe the order in which the animals in the different experimental groups were treated and assessed.
Experimental outcomes	12	Clearly define the primary and secondary experimental outcomes assessed (e.g., cell death, molecular markers, behavioural changes).
Statistical methods	13	a. Provide details of the statistical methods used for each analysis. b. Specify the unit of analysis for each dataset (e.g. single animal, group of animals, single neuron). c. Describe any methods used to assess whether the data met the assumptions of the statistical approach.
RESULTS		
Baseline data	14	For each experimental group, report relevant characteristics and health status of animals (e.g., weight, microbiological status, and drug- or test-naïve) before treatment or testing (this information can often be tabulated).
Numbers analysed	15	a. Report the number of animals in each group included in each analysis. Report absolute numbers (e.g. 10/20, not 50%[a]). b. If any animals or data were not included in the analysis, explain why.
Outcomes and estimation	16	Report the results for each analysis carried out, with a measure of precision (e.g., standard error or confidence interval).
Adverse events	17	a. Give details of all important adverse events in each experimental group. b. Describe any modifications to the experimental protocols made to reduce adverse events.

Table 2. Cont.

	ITEM	RECOMMENDATION
DISCUSSION		
Interpretation/scientific implications	18	a. Interpret the results, taking into account the study objectives and hypotheses, current theory, and other relevant studies in the literature. b. Comment on the study limitations including any potential sources of bias, any limitations of the animal model, and the imprecision associated with the results[a]. c. Describe any implications of your experimental methods or findings for the replacement, refinement, or reduction (the 3Rs) of the use of animals in research.
Generalisability/translation	19	Comment on whether, and how, the findings of this study are likely to translate to other species or systems, including any relevance to human biology.
Funding	20	List all funding sources (including grant number) and the role of the funder(s) in the study.

[a]Schulz, et al. (2010) [24].

doi:10.1371/journal.pbio.1000412.t002

References

1. Simera I, Altman DG (2009) Writing a research article that is "fit for purpose": EQUATOR Network and reporting guidelines. Evid Based Med 14: 132–134.
2. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, et al. (2010) Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med 7: e1000245. doi:10.1371/journal.pmed.1000245.
3. Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR (2010) Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. PLoS Biol 8: e1000344. doi:10.1371/journal.pbio.1000344.
4. Sargeant JM, Thompson A, Valcour J, Elgie R, Saint-Onge J, et al. (2010) Quality of reporting of clinical trials of dogs and cats and associations with treatment effects. J Vet Intern Med 24: 44–50.
5. Kilkenny C, Parsons N, Kadyszewski E, Festing MFW, Cuthill IC, et al. (2009) Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. PLoS ONE 4: e7824. doi:10.1371/journal.pone.0007824.
6. Sargeant JM, Elgie R, Valcour J, Saint-Onge J, Thompson A, et al. (2009) Methodological quality and completeness of reporting in clinical trials conducted in livestock species. Prev Vet Med 91: 107–115.
7. Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, et al. (2009) Good laboratory practice. Preventing introduction of bias at the bench. Stroke 40: 50–52.
8. Hainsworth AH, Markus HS (2008) Do *in vivo* experimental models reflect human cerebral small vessel disease? A systematic review. J Cereb Blood Flow Metab 28: 1877–1891.
9. Rice ASC, Cimino-Brown D, Eisenach JC, Kontinen VK, Lacroix-Fralish ML, et al. (2008) Animal models and the prediction of efficacy in clinical trials of analgesic drugs: a critical appraisal and call for uniform reporting standards. Pain 139: 243–247.
10. Sherwin CM (2007) Animal welfare: reporting details is good science. Nature 448: 251.
11. Jafari P, Azuaje F (2006) An assessment of recently published gene expression analyses: reporting experimental design and statistics. BMC Med Inform Decis Mak 6: 27.
12. Hackam DG, Redelmeier DA (2006) Translation of research evidence from animals to humans. JAMA 296: 1731–1732.
13. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, et al. (2006) Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ 334: 197. doi:10.1136/bmj.39048.407928.BE.
14. Macleod M (2005) What can systematic review and meta-analysis tell us about the experimental data supporting stroke drug development? Int J Neuroprot Neuroregener 1: 201.
15. Tooth L, Ware R, Bain C, Purdie DM, Dobson A (2005) Quality of reporting of observational longitudinal research. Am J Epidemiol 161: 280–288.
16. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I (2004) Where is the evidence that animal research benefits humans? BMJ 328: 514–517.
17. Bennett LT, Adams M A (2004) Assessment of ecological effects due to forest harvesting: approaches and statistical issues. J Appl Ecol 41: 585–598.
18. Morris CE, Bardin M, Berge O, Frey-Klett P, Fromin N (2002) Microbial biodiversity: Approaches to experimental design and hypothesis testing in primary scientific literature from 1975 to 1999. Microbiol Mol Biol Rev 66: 592–616.
19. Smith JA, Birke L, Sadler D (1997) Reporting animal use in scientific papers. Lab Anim 31: 312–317.
20. McCance I (1995) Assessment of statistical procedures used in papers in the Australian Veterinary Journal. Aust Vet J 72: 322–328.
21. Pocock SJ, Hughes MD, Lee RJ (1987) Statistical problems in the reporting of clinical trials. A survey of three medical journals. New Engl J Med 317: 426–432.
22. Chalmers I, Glasziou P (2009) Avoidable waste in the production and reporting of research evidence. Lancet 374: 86–89.
23. Festing MF, Altman DG (2002) Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR J 43: 244–258.
24. Schulz KF, Altman DG, Moher D, the CONSORT Group (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 340: c332.
25. Moher D, Schulz KF, Altman DG for the CONSORT Group (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357: 1191–1194.
26. Altman DG (2005) Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. BMJ 330: 1056–1057.
27. Hopewell S, Altman DG, Moher D, Schulz KF (2008) Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'instructions to Authors'. Trials 9: 20.
28. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, et al. (2006) Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust 185: 263–267.
29. Kane RL, Wang J, Garrard J (2007) Reporting in randomized clinical trials improved after adoption of the CONSORT statement. J Clin Epidemiol 60: 241–249.
30. Altman DG, Simera I, Hoey J, Moher D, Schulz K (2008) EQUATOR: reporting guidelines for health research. Lancet 371: 1149–1150.
31. Simera I, Moher D, Hoey J, Schulz K, Altman DG (2010) A catalogue of reporting guidelines for health research. Eur J Clin Invest 40: 35–53.
32. Wager E, Field EA, Grossman L (2003) Good publication practice for pharmaceutical companies. Curr Med Res Opin 19: 149–154.
33. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. (2001) Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. Nat Genet 29: 365–371.
34. Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker DJ, et al. (2007) Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 3: 231–241.
35. Stone SP, Cooper BS, Kibbler CC, Cookson BD, Roberts JA (2007) The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. J Antimicrob Chemother 59: 833–840.
36. Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KR (2006) A systematic review of systematic reviews and meta-analyses of animal experiments with guidelines for reporting. J Environ Sci Health B 41: 1245–1258.
37. O'Connor AM, Sargeant JM, Gardner IA, Dickson JS, Torrence ME, et al. (2010) The REFLECT statement: Methods and processes of creating reporting guidelines for randomised controlled trials for livestock and food safety. Prev Vet Med 93: 11–18.
38. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals. Available: http://www.icmje.org/urm_full.pdf. Accessed 22 January 2010.
39. Council of Science Editors – Editorial Policy Committee (2008–2009) CSE's White Paper on promoting integrity in scientific journal publications, 2009 Update. Available: http://www.councilscienceeditors.org/editorial_policies/whitepaper/entire_whitepaper.pdf. Accessed 22 January 2010.

40. Committee on publication ethics (COPE). COPE best practice guidelines for journal editors. Available: http://publicationethics.org/files/u2/Best_Practice.pdf. Accessed 22 January 2010.
41. Nuffield Council on Bioethics: The Ethics of Research involving Animals (2005) Chapter 15: Discussion and Recommendations; pp 313, paragraph 15.58. Available: http://www.nuffieldbioethics.org/fileLibrary/pdf/RIA_Report_FINAL-opt.pdf. Accessed 26 January 2010.
42. Drummond GB (2009) Reporting ethical matters in The Journal of Physiology: standards and advice. J Physiol 587.4: 713–719.
43. Osborne NJ, Payne D, Newman ML (2009) Journal editorial policies, animal welfare, and the 3Rs. Am J Bioeth 9: 55–59.
44. Hooijmans C, Leenars M, Ritskes-Hoitinga M (2009) Improving the quality of publications on animal experiments to make systematic reviews possible. ALTEX 26: 262.
45. Wurbel H (2007) Publications should include an animal welfare section. Nature 446: 257.
46. Alfaro V (2005) Specification of laboratory animal use in scientific articles: Current low detail in the journal's instructions for authors and some proposals. Methods Find Exp Clin Pharmacol 27: 495–502.
47. Nuffield Council on Bioethics (2005) The ethics of research involving animals. London: Ed Nuffield Council on Bioethics, Available: http://www.nuffieldbioethics.org/go/browseablepublications/ethicsofresearchanimals/report_490.html. Accessed 22 January 2010.
48. Öbrink KJ, Rehbinder C (2000) Animal definition: a necessity for the validity of animal experiments? Lab Anim 34:: 1 21–130.
49. Boisvert DPJ (1997) Editorial policies and animal welfare. In Animal Alternatives, Welfare and Ethics Zutphen LFM, Balls M, eds. Elsevier. pp 399–404.
50. Working Committee for the Biological characterisation of laboratory animals/GV-Solas (1985) Guidelines for specification of animals and husbandry methods when reporting the results of animal experiments. Lab Anim 19: 106–108.
51. Moher D, Schulz K, Simera I, Altman DG (2010) Guidance for developers of health research reporting guidelines. PLoS Med 7: e1000217. doi:10.1371/journal.pmed.1000217.
52. Mignini LE, Khan KS (2006) Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research. Biomed Central Medical Research Methodology 6: 10.
53. Macleod MR, Collins T, Howells DW, Donnan GA (2004) Pooling of animal experimental data reveals influence of study design and publication bias. Stroke 35: 1203–1208.
54. Roberts I, Kwan I, Evans P, Haig S (2002) Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. BMJ 324: 474–476.
55. Horn J, de Haan RJ, Vermeulen M, Limburg M (2001) Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. Stroke 32: 2433–38.
56. Kilkenny C, Brown WJ, Cuthill IC, Emerson M, Altman DG (2010) Animal Research: Reporting *In Vivo* Experiments: The ARRIVE guidelines. J Gene Med. doi: 10.1002/jgm.1473.
57. Kilkenny C, Brown WJ, Cuthill IC, Emerson M, Altman DG (2010) Animal Research: Reporting *In Vivo* Experiments: The ARRIVE guidelines. Exp Physiol. doi: 10.1113/expphysiol.2010.053793.
58. Kilkenny C, Brown WJ, Cuthill IC, Emerson M, Altman DG (2010) Animal Research: Reporting *In Vivo* Experiments: The ARRIVE guidelines. J Physiol. doi: 10.1113/jphysiol.2010.192278.
59. Kilkenny C, Brown WJ, Cuthill IC, Emerson M, Altman DG (2010) Animal Research: Reporting *In Vivo* Experiments: The ARRIVE guidelines. Br J Pharmacol. doi:10.1111/j.1476-5381.2010.00872.x.
60. Kilkenny C, Brown WJ, Cuthill IC, Emerson M, Altman DG (2010) Animal Research: Reporting *In Vivo* Experiments: The ARRIVE guidelines. Laboratory Animals. doi:10.1258/la.2010.0010021.

The ARRIVE guidelines 2019: updated guidelines for reporting animal research

Nathalie Percie du Sert[1], Viki Hurst[1], Amrita Ahluwalia[2], Sabina Alam[3], Marc T. Avey[4], Monya Baker[5], William J. Browne[6], Alejandra Clark[7], Innes C. Cuthill[6], Ulrich Dirnagl[8], Michael Emerson[9], Paul Garner[10], Stephen T. Holgate[11], David W. Howells[12], Natasha A. Karp[13], Katie Lidster[1], Catriona J. MacCallum[14], Malcolm Macleod[15], Ole Petersen[16], Frances Rawle[17], Penny Reynolds[18], Kieron Rooney[19], Emily S. Sena[15], Shai D. Silberberg[20], Thomas Steckler[21], Hanno Würbel[22]

1 NC3Rs, UK
2 Queen Mary University of London, UK
3 Taylor & Francis Group, UK
4 ICF, USA
5 Nature, USA
6 University of Bristol, UK
7 PLOS ONE, UK
8 Universitätsmedizin Berlin, Germany
9 Imperial College London, UK
10 Liverpool School of Tropical Medicine, UK
11 University of Southampton, UK
12 University of Tasmania, Australia
13 Data Sciences & Quantitative Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK
14 Hindawi Ltd, UK
15 Centre for Clinical Brain Sciences, University of Edinburgh, UK
16 Cardiff University, UK
17 Medical Research Council, UK
18 University of Florida, USA
19 University of Sydney, Australia
20 National Institute of Neurological Disorders and Stroke, USA
21 Janssen Pharmaceutica NV, Belgium
22 University of Bern, Switzerland

Abstract

Reproducible science requires transparent reporting. The ARRIVE guidelines were originally developed in 2010 to improve the reporting of animal research. They consist of a checklist of information to include in publications describing *in vivo* experiments to enable others to scrutinise the work adequately, evaluate its methodological rigour, and reproduce the methods and results. Despite considerable levels of endorsement by funders and journals over the years, adherence to the guidelines has been inconsistent, and the anticipated improvements in the quality of reporting in animal research publications have not been achieved.

Here we introduce ARRIVE 2019. The guidelines have been updated and information reorganised to facilitate their use in practice. We used a Delphi exercise to prioritise the items and split the guidelines into two sets, the ARRIVE Essential 10, which constitute the minimum requirement, and the Recommended Set, which describes the research context. This division facilitates improved reporting of animal research by supporting a stepwise approach to implementation. This helps journal editors and reviewers to verify that the most important items are being reported in manuscripts. We have also developed the accompanying Explanation and Elaboration document that serves 1) to explain the rationale behind each item in the guidelines, 2) to clarify key concepts and 3) to provide illustrative examples. We aim through these changes to help ensure that researchers, reviewers and journal editors are better equipped to improve the rigour and transparency of the scientific process and thus reproducibility.

Why good reporting is important

In recent years issues about the reproducibility of research findings have raised considerable concern among scientists, funders, research users and policy makers [1-3]. Important contributing factors identified include flawed study design and analysis, variability and inadequate validation of reagents and other biological materials, insufficient reporting of methodology and results, and barriers to access data [4]. A number of initiatives have been developed to improve the reproducibility of scientific research, from funders' open access policies [5], through to alternative peer review models [6], and the development of infrastructure to promote study preregistration and data sharing [7].

Transparent reporting is an essential first step for any initiative focusing on reproducibility. Without this, the methodological rigour of the studies cannot be adequately scrutinised, the reliability of the findings cannot be assessed, and the work cannot be repeated or built upon by others. Despite the development of specific reporting guidelines for preclinical and clinical research, evidence suggests that scientific publications often lack key information and that there continues to be considerable scope for improvement [8-14]. Animal research is a good case in point, where poor reporting impacts on the development of therapeutics and irreproducible publications can spawn an entire field of research, or trigger clinical studies, subjecting patients to interventions unlikely to be effective [2, 15, 16].

In an attempt to improve the reporting of animal research, the ARRIVE guidelines were published in 2010. The guidelines consist of a checklist of the information that should be included in any manuscript describing animal-based research, to ensure that the research is described in a comprehensive and transparent manner [17-27]. In the nine years since publication, the ARRIVE guidelines have been endorsed by more than a thousand journals from across the life sciences. Endorsement typically includes advocating their use in guidance to authors and reviewers. However, only a small number of journals actively enforce compliance; recent studies have shown that important information as set out in the ARRIVE guidelines is still missing from most publications sampled. This includes randomisation (reported in only 30-40% of publications), blinding (reported in only approximately 20% of publications), sample size justification (reported in less than 10% of publications) and animal characteristics (all basic characteristics reported in less than 10% of publications) [28-30].

Evidence suggests two main factors limit the impact of the guidelines. The first is the extent to which editorial and journal staff are involved in enforcing reporting standards. A randomised controlled trial at PLOS ONE, for example, demonstrated that a request by journal staff to include a completed ARRIVE checklist in the manuscript submission process did not improve the disclosure of information in published papers [31]. In contrast, other studies using reporting checklists with more editorial follow up have shown a marked improvement in the nature and detail of the information included in publications [32-34]. Providing the level of journal or editorial input required to ensure compliance with all the items of the ARRIVE guidelines is unlikely to be sustainable for most journals because of the resources needed. Requesting adherence with all items at once, with no consideration of their relative importance might also be perceived as too prescriptive and further complicate the task.

The second issue is that researchers and other individuals and organisations responsible for the integrity of the research process are not sufficiently aware of the consequences of incomplete reporting. There is some evidence that awareness of ARRIVE is linked to the use of more rigorous experimental design standards [35], but there is also evidence that researchers are unaware of the much larger systemic bias in the publication of research and in the reliability of certain findings and even of entire fields [31, 36-38]. This lack of understanding affects how experiments are designed and grant proposals prepared, how animals are handled and data recorded in the laboratory, and how manuscripts are written by authors or assessed by journal staff, editors and reviewers.

Approval for experiments involving animals is generally based on a harm-benefit analysis, weighing the harms to the animals involved against the benefits of the research to society. If the research is not reported in enough detail, even when conducted rigorously, the benefits may not be realised, and the harm-benefit analysis and public trust in the research are undermined [39]. As a community, we must do better to ensure that where animals are used the research is well designed and analysed, and transparently reported. Here we introduce the revised ARRIVE guidelines, referred to as ARRIVE 2019. The information included has been updated, extended and reorganised to facilitate the use of

the guidelines, helping to ensure that researchers, editors and reviewers, as well as other relevant journal staff, are better equipped to improve the rigour and reproducibility of animal research.

Introducing ARRIVE 2019

The revision of the ARRIVE guidelines has been undertaken by a new international working group – the authors of this publication. Our expertise comes from across the life sciences community, including funders, journal editors, statisticians, methodologists and researchers from academia and industry. We have improved the clarity of the guidelines, prioritised the items, added new information and generated the accompanying Explanation and Elaboration document to provide context and rationale for each item [40]. New additions comprise inclusion and exclusion criteria, which are a key aspect of data handling and prevent the ad hoc exclusion of data [41]; protocol registration, a recently emerged approach which promotes scientific rigour and encourages researchers to carefully consider the experimental design and analysis plan before any data are collected [42]; and data access, in line with the FAIR Data Principles [43]. Table S1 summarises the changes.

The most significant departure from the original guidelines is the classification of items into two prioritised groups, as shown in Tables 1 and 2. There is no ranking within each group. The first group is the “ARRIVE Essential 10” which describes information that is the basic minimum to include in a manuscript, as without this information reviewers and readers cannot confidently assess the reliability of the findings presented. It includes the study design, sample size, measures to reduce subjective bias such as randomisation and blinding, outcome measures, statistical methods, experimental animals used, experimental procedures and results. The second group, referred to as the “Recommended Set” adds context to the study described. This includes the ethical statement, declaration of interest, protocol registration and data access, as well as more detailed information on the methodology such as animal housing, husbandry, care and monitoring. Items on abstract, background, objectives, interpretation and generalisability also describe what to include in the more narrative parts of a manuscript. The prioritisation was derived from a Delphi exercise [44] to rank items according to their relative importance for assessing the reliability of research findings. The Delphi panel included the working group in addition to external stakeholders, to ensure maximum diversity in fields of expertise and geographical location. Demographics of the Delphi panel and full methods and results are presented in Supporting Information S2 and S3.

The classification of the items into two groups is intended to facilitate the improved reporting of animal research by allowing an initial focus on the most critical issues. This better allows journal staff, editors and reviewers to verify that the items have been adequately reported in manuscripts. The first step should be to ensure compliance with the ARRIVE Essential 10 as a minimum requirement. Items from the Recommended Set can then be added over time and in line with specific editorial policies until all the items are routinely reported in all manuscripts.

Although the guidelines are written with researchers and journal editorial policies in mind, it is important to stress that researchers alone should not have to carry the burden of responsibility for transparent reporting. Funders, institutions and publishers all have a responsibility to ensure that the appropriate training, workflows and practices are in place to support researchers in their different roles. In particular, institutions and other research performing organisations, both public and private, as well as publishers, have a key role to play in building capacity and championing the behavioural changes required to improve reporting practices.

ARRIVE Essential 10		
Study design	1	For each experiment, provide brief details of study design including: a. The groups being compared, including control groups. If no control group has been used, the rationale should be stated. b. The experimental unit (e.g. a single animal, litter, or cage of animals).
Sample size	2	a. Specify the exact number of experimental units allocated to each group, and the total number in each experiment. Also indicate the total number of animals used. b. Explain how the sample size was decided. Provide details of any *a priori* sample size calculation, if done.
Inclusion and exclusion criteria	3	a. Describe any criteria established *a priori* for including and excluding animals (or experimental units) during the experiment, and data points during the analysis. b. For each experimental group, report any animals, experimental units or data points not included in the analysis and explain why. c. For each analysis, report the exact value of N in each experimental group.
Randomisation	4	Describe the methods used: a. To allocate experimental units to control and treatment groups. If randomisation was used, provide the method of randomisation. b. To minimise potential confounding factors such as the order of treatments and measurements, or animal/cage location.
Blinding	5	Describe who was aware of the group allocation at the different stages of the experiment (during the allocation, the conduct of the experiment, the outcome assessment, and the data analysis).
Outcome measures	6	a. Clearly define all outcome measures assessed (e.g. cell death, molecular markers, or behavioural changes). b. For hypothesis-testing studies, specify the primary outcome measure, i.e. the outcome measure that was used to determine the sample size.
Statistical methods	7	a. Provide details of the statistical methods used for each analysis. b. Specify the experimental unit that was used for each statistical test. c. Describe any methods used to assess whether the data met the assumptions of the statistical approach.
Experimental animals	8	a. Provide details of the animals used, including species, strain and substrain, sex, age or developmental stage, and weight. b. Provide further relevant information on the provenance of animals, health/immune status, genetic modification status, genotype, and any previous procedures.
Experimental procedures	9	For each experimental group, including controls, describe the procedures in enough detail to allow others to replicate them, including: a. What was done, how it was done and what was used. b. When and how often. c. Where (including detail of any acclimation periods). d. Why (provide rationale for procedures).
Results	10	For each experiment conducted, including independent replications, report: a. Summary/descriptive statistics for each experimental group, with a measure of variability where applicable. b. If applicable, the effect size with a confidence interval.

Table 1. ARRIVE Essential 10

Recommended Set		
Abstract	11	Provide an accurate summary of the research objectives, animal species, strain and sex, key methods, principal findings, and study conclusions.
Background	12	a. Include sufficient scientific background to understand the rationale and context for the study, and explain the experimental approach. b. Explain how the animal species and model used address the scientific objectives and, where appropriate, the relevance to human biology.
Objectives	13	Clearly describe the research question, research objectives and, where appropriate, specific hypotheses being tested.
Ethical statement	14	Provide the name of the ethical review committee or equivalent that has approved the use of animals in this study and any relevant licence or protocol numbers (if applicable). If ethical approval was not sought or granted, provide a justification.
Housing and husbandry	15	Provide details of housing and husbandry conditions, including any environmental enrichment.
Animal care and monitoring	16	a. Describe any interventions or steps taken in the experimental protocols to reduce pain, suffering and distress. b. Report any expected or unexpected adverse events. c. Describe the humane endpoints established for the study and the frequency of monitoring.
Interpretation /scientific implications	17	a. Interpret the results, taking into account the study objectives and hypotheses, current theory and other relevant studies in the literature. b. Comment on the study limitations including potential sources of bias, limitations of the animal model, and imprecision associated with the results.
Generalisability /translation	18	Comment on whether, and how, the findings of this study are likely to generalise to other species or experimental conditions, including any relevance to human biology (where appropriate).
Protocol registration	19	Provide a statement indicating whether a protocol (including the research question, key design features, and analysis plan) was prepared before the study, and if and where this protocol was registered.
Data access	20	Provide a statement describing if and where study data are available.
Declaration of interests	21	a. Declare any potential conflicts of interest, including financial and non-financial. If none exist, this should be stated. b. List all funding sources (including grant identifier) and the role of the funder(s) in the design, analysis and reporting of the study.

Table 2. ARRIVE Recommended Set

Conclusion

Transparent reporting is clearly essential if animal studies are to add to the knowledge base and inform future research, policy and clinical practice. ARRIVE 2019 prioritises the reporting of information related to study reliability. This enables research users to assess how much weight to ascribe to the findings, and in parallel promotes the use of rigorous methodology in the planning and conduct of *in vivo* experiments [35], thus increasing the likelihood that the findings are reliable, and ultimately, reproducible.

The intention of ARRIVE 2019 is not to supersede individual journal requirements but to promote a harmonised approach across journals to ensure that all manuscripts contain essential information needed to appraise the research. The step-by-step approach is in line with current practice by a number of journals who recommend that authors not only refer to the ARRIVE guidelines while preparing their manuscript, but also implement a checklist with a core set of items. Journals usually share a common objective of improving the methodological rigour and reproducibility of the research they publish, but different journals emphasise different pieces of information [45-47]. Here we propose

an expert consensus on information to prioritise. This will provide clarity for authors, facilitate transfer of manuscripts between journals, and accelerate an improvement of reporting standards.

Concentrating the efforts of the research and publishing communities on the ARRIVE Essential 10 items provides a manageable approach to evaluate reporting quality efficiently and assess the effect of interventions and policies designed to improve the reporting of animal experiments. It also provides a starting point for the development of automated or semi-automated artificial intelligence tools that can detect missing information rapidly [48].

Improving reporting is a collaborative endeavour and concerted effort from the biomedical research community is required to ensure maximum impact. We welcome collaboration with other groups operating in this area, and feedback on ARRIVE 2019 and our implementation strategy.

Supporting information

S1 Table: Noteworthy changes in ARRIVE 2019, compared to ARRIVE 2010

S2 Delphi methods and results

S3 Delphi data

Acknowledgements

We would like to thank the members of the expert panel for the Delphi exercise, and the DelphiManager team for advice and use of their software. We would like to acknowledge the late Doug Altman's contribution to this project, Doug was a dedicated member of the working group and his input to the guidelines' revision has been invaluable.

Competing interests

AA: editor in chief of the British Journal of Pharmacology. WJB, ICC and ME: authors of the original ARRIVE guidelines. WJB: serves on the Independent Statistical Standing Committee of the funder CHDI foundation. AC, CJM, MMcL and ESS: involved in the IICARus trial. ME, MMcL and ESS: have received funding from NC3Rs. ME: sits on the MRC ERPIC panel. STH: chair of the NC3Rs board, trusteeship of the BLF, Kennedy Trust, DSRU and CRUK, member of Governing Board, Nuffield Council of Bioethics, member Science Panel for Health (EU H2020), founder and NEB Director Synairgen, consultant Novartis, Teva and AZ, chair MRC/GSK EMINENT Collaboration. KL, VH and NPdS: NC3Rs staff, role includes promoting the ARRIVE guidelines. CJMcC: shareholdings in Hindawi, on the publishing board of the Royal Society, on the EU Open Science policy platform. MMcL, NPdS, CJMcC, ESS, TS and HW: members of EQIPD. MMcL: member of the Animals in Science Committee. NPdS and TS: associate editors of BMJ Open Science. OP: vice president of Academia Europaea, senior executive editor of the Journal of Physiology, member of the Board of the European Commission's SAPEA (Science Advice for Policy by European Academies). FR: NC3Rs board member, shareholdings in AstraZeneca and GSK. PR: member of the University of Florida Institutional Animal Care and Use Committee, Editorial board member of Shock. ESS: editor in chief of BMJ Open Science. SDS: role is to provide expertise and does not represent the opinion of the NIH. TS: shareholdings in Johnson & Johnson. SA, MTA, MB, UD, PG, DWH, NAK and KR declared no conflict of interest.

References

1. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? Sci Transl Med. 2016;8(341):341ps12. doi: 10.1126/scitranslmed.aaf5027.

2. Begley CG, Ioannidis JP. Reproducibility in science: Improving the standard for basic and preclinical research. Circ Res. 2015;116(1):116-26. Epub 2015/01/02. doi: 10.1161/CIRCRESAHA.114.303819. PubMed PMID: 25552691.

3. Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016;533(7604):452-4. Epub 2016/05/27. doi: 10.1038/533452a. PubMed PMID: 27225100.

4. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: Progress and priorities. F1000Res. 2017;6:604. Epub 2017/06/18. doi: 10.12688/f1000research.11334.1. PubMed PMID: 28620458; PubMed Central PMCID: PMCPMC5461896.

5. Else H. Radical open-access plan could spell end to journal subscriptions. Nature. 2018;561(7721):17-8. Epub 2018/09/06. doi: 10.1038/d41586-018-06178-7. PubMed PMID: 30181639.

6. Chambers CD, Forstmann B, Pruszynski JA. Registered reports at the european journal of neuroscience: Consolidating and extending peer-reviewed study pre-registration. Eur J Neurosci. 2017;45(5):627-8. Epub 2016/12/28. doi: 10.1111/ejn.13519. PubMed PMID: 28027598.

7. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci U S A. 2018;115(11):2600-6. Epub 2018/03/14. doi: 10.1073/pnas.1708274114. PubMed PMID: 29531091; PubMed Central PMCID: PMCPMC5856500.

8. Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, et al. Good laboratory practice: Preventing introduction of bias at the bench. Stroke. 2009;40(3):e50-2. Epub 2008/08/16. doi: STROKEAHA.108.525386 [pii] 10.1161/STROKEAHA.108.525386. PubMed PMID: 18703798.

9. Rice AS, Cimino-Brown D, Eisenach JC, Kontinen VK, Lacroix-Fralish ML, Machin I, et al. Animal models and the prediction of efficacy in clinical trials of analgesic drugs: A critical appraisal and call for uniform reporting standards. Pain. 2009;139(2):243-7. Epub 2008/09/26. doi: S0304-3959(08)00508-3 [pii] 10.1016/j.pain.2008.08.017. PubMed PMID: 18814968.

10. McCance I. Assessment of statistical procedures used in papers in the australian veterinary journal. Aust Vet J. 1995;72(9):322-8. Epub 1995/09/01. PubMed PMID: 8585846.

11. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. JAMA. 2006;296(14):1731-2. Epub 2006/10/13. doi: 10.1001/jama.296.14.1731. PubMed PMID: 17032985.

12. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One. 2009;4(11):e7824. Epub 2009/12/04. doi: 10.1371/journal.pone.0007824. PubMed PMID: 19956596.

13. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? PLoS Med. 2010;7(3):e1000245. Epub 2010/04/03. doi: 10.1371/journal.pmed.1000245. PubMed PMID: 20361020.

14. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383(9913):267-76. Epub 2014/01/15. doi: 10.1016/S0140-6736(13)62228-X. PubMed PMID: 24411647.

15. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483(7391):531-3. Epub 2012/03/31. doi: 10.1038/483531a. PubMed PMID: 22460880.

16. Scott S, Kranz JE, Cole J, Lincecum JM, Thompson K, Kelly N, et al. Design, power, and interpretation of studies in the standard murine model of als. Amyotroph Lateral Scler. 2008;9(1):4-15. Epub 2008/02/15. doi: 789666722 [pii] 10.1080/17482960701856300. PubMed PMID: 18273714.

17. Kilkenny C, Altman DG. Improving bioscience research reporting: ARRIVE-ing at a solution. Lab Anim. 2010;44(4):377-8. Epub 2010/07/28. doi: la.2010.0010021 [pii] 10.1258/la.2010.0010021. PubMed PMID: 20660161.

18. Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG, Group NCRRGW. Animal research: Reporting in vivo experiments: The ARRIVE guidelines. Br J Pharmacol. 2010;160(7):1577-9. Epub 2010/07/24. doi: 10.1111/j.1476-5381.2010.00872.x. PubMed PMID: 20649561; PubMed Central PMCID: PMCPMC2936830.

19. Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG, Group NCRRGW. Animal research: Reporting in vivo experiments: The ARRIVE guidelines. J Gene Med. 2010;12(7):561-3. Epub 2010/07/08. doi: 10.1002/jgm.1473. PubMed PMID: 20607692.

20. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLoS Biol. 2010;8(6):e1000412. Epub 2010/07/09. doi: 10.1371/journal.pbio.1000412. PubMed PMID: 20613859; PubMed Central PMCID: PMC2893951.

21. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. Journal of Pharmacology & Pharmacotherapeutics. 2010;1(2):94-9. doi: 10.4103/0976-500X.72351.

22. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Animal research: Reporting in vivo experiments: The ARRIVE guidelines. J Physiol. 2010;588(Pt 14):2519-21. Epub 2010/07/17. doi: 588/14/2519 [pii] 10.1113/jphysiol.2010.192278. PubMed PMID: 20634180; PubMed Central PMCID: PMC2916981.

23. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Animal research: Reporting in vivo experiments: The ARRIVE guidelines. Exp Physiol. 2010;95(8):842-4. Epub 2010/07/09. doi: 95/8/842 [pii] 10.1113/expphysiol.2010.053793. PubMed PMID: 20610776.

24. McGrath JC, Drummond GB, McLachlan EM, Kilkenny C, Wainwright CL. Guidelines for reporting experiments involving animals: The ARRIVE guidelines. Br J Pharmacol. 2010;160(7):1573-6. Epub 2010/07/24. doi: BPH873 [pii] 10.1111/j.1476-5381.2010.00873.x. PubMed PMID: 20649560; PubMed Central PMCID: PMC2936829.

25. Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG. Animal research: Reporting in vivo experiments-the ARRIVE guidelines. J Cereb Blood Flow Metab. 2011. Epub 2011/01/06. doi: jcbfm2010220 [pii] 10.1038/jcbfm.2010.220. PubMed PMID: 21206507.

26. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. Vet Clin Pathol. 2012;41(1):27-31. Epub 2012/03/07. doi: 10.1111/j.1939-165X.2012.00418.x. PubMed PMID: 22390425.

27. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. Osteoarthritis Cartilage. 2012;20(4):256-60. doi: 10.1111/j.1939-165X.2012.00418.x.

28. Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, et al. The devil is in the details: Incomplete reporting in preclinical animal research. PLoS One. 2016;11(11):e0166733.

Epub 2016/11/18. doi: 10.1371/journal.pone.0166733. PubMed PMID: 27855228; PubMed Central PMCID: PMCPMC5113978.

29. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of bias in reports of in vivo research: A focus for improvement. PLoS Biol. 2015;13(10):e1002273. Epub 2015/10/16. doi: 10.1371/journal.pbio.1002273. PubMed PMID: 26460723; PubMed Central PMCID: PMCPMC4603955.

30. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: Support for the ARRIVE (animal research: Reporting of in vivo experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. PLoS One. 2018;13(5):e0197882. Epub 2018/05/26. doi: 10.1371/journal.pone.0197882. PubMed PMID: 29795636; PubMed Central PMCID: PMCPMC5967836.

31. Hair K, Macleod MR, Sena ES, Sena ES, Hair K, Macleod MR, et al. A randomised controlled trial of an intervention to improve compliance with the ARRIVE guidelines (iicarus). Research Integrity and Peer Review. 2019;4(1):12. doi: 10.1186/s41073-019-0069-3.

32. The NPQIP Collaborative group. Did a change in nature journals' editorial policy for life sciences research improve reporting? BMJ Open Science. 2019;3(1):e000035. doi: 10.1136/bmjos-2017-000035.

33. Han S, Olonisakin TF, Pribis JP, Zupetic J, Yoon JH, Holleran KM, et al. A checklist is associated with increased quality of reporting preclinical biomedical research: A systematic review. PLoS One. 2017;12(9):e0183591. Epub 2017/09/14. doi: 10.1371/journal.pone.0183591. PubMed PMID: 28902887; PubMed Central PMCID: PMCPMC5597130.

34. Ramirez FD, Motazedian P, Jung RG, Di Santo P, MacDonald ZD, Moreland R, et al. Methodological rigor in preclinical cardiovascular studies: Targets to enhance reproducibility and promote research translation. Circ Res. 2017;120(12):1916-26. Epub 2017/04/05. doi: 10.1161/CIRCRESAHA.117.310628. PubMed PMID: 28373349; PubMed Central PMCID: PMCPMC5466021.

35. Reichlin TS, Vogt L, Wurbel H. The researchers' view of scientific rigor-survey on the conduct and reporting of in vivo research. PLoS One. 2016;11(12):e0165999. Epub 2016/12/03. doi: 10.1371/journal.pone.0165999. PubMed PMID: 27911901; PubMed Central PMCID: PMCPMC5135049.

36. Hurst V, Percie du Sert N. The ARRIVE guidelines survey. Open Science Framework. 2017. doi: https://osf.io/g8t5q/register/565fb3678c5e4a66b5582f67.

37. Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. Questionable research practices in ecology and evolution. PLoS One. 2018;13(7):e0200303. Epub 2018/07/17. doi: 10.1371/journal.pone.0200303. PubMed PMID: 30011289; PubMed Central PMCID: PMCPMC6047784.

38. The Academy of Medical Sciences. Reproducibility and reliability of biomedical research: Improving research practice 2015. Available from: https://acmedsci.ac.uk/policy/policy-projects/reproducibility-and-reliability-of-biomedical-research.

39. Sena ES, Currie GL. How our approaches to assessing benefits and harms can be improved. Animal Welfare. 2019;28(1):107-15. doi: 10.7120/09627286.28.1.107. PubMed PMID: WOS:000455904200011.

40. Percie du Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, et al. Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2019. Submitted to Plos Biology. 2019.

41. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012;490(7419):187-91. Epub 2012/10/13. doi: 10.1038/nature11556. PubMed PMID: 23060188; PubMed Central PMCID: PMC3511845.

42. Kimmelman J, Anderson JA. Should preclinical studies be registered? Nat Biotechnol. 2012;30(6):488-9. Epub 2012/06/09. doi: 10.1038/nbt.2261. PubMed PMID: 22678379.

43. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The fair guiding principles for scientific data management and stewardship. Scientific Data. 2016;3:160018. doi: 10.1038/sdata.2016.18.

44. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7(2):e1000217. Epub 2010/02/20. doi: 10.1371/journal.pmed.1000217. PubMed PMID: 20169112.

45. Prager EM, Chambers KE, Plotkin JL, McArthur DL, Bandrowski AE, Bansal N, et al. Improving transparency and scientific rigor in academic publishing. J Neurosci Res. 2018. Epub 2018/12/07. doi: 10.1002/jnr.24340. PubMed PMID: 30506706.

46. Enhancing reproducibility. Nat Methods. 2013;10(5):367. Epub 2013/06/14. PubMed PMID: 23762900.

47. Curtis MJ, Alexander S, Cirino G, Docherty JR, George CH, Giembycz MA, et al. Experimental design and analysis and their reporting ii: Updated and simplified guidance for authors and peer reviewers. Br J Pharmacol. 2018;175(7):987-93. Epub 2018/03/10. doi: 10.1111/bph.14153. PubMed PMID: 29520785; PubMed Central PMCID: PMCPMC5843711.

48. Heaven D. Ai peer reviewers unleashed to ease publishing grind. Nature. 2018;563(7733):609-10. Epub 2018/11/30. doi: 10.1038/d41586-018-07245-9. PubMed PMID: 30482927.

S1 Table: Noteworthy changes in ARRIVE 2019, compared to ARRIVE 2010

ARRIVE 2019	ARRIVE 2010	Reason for change
All items	All items	We reordered items and split them in two sets based on their importance to assess the reliability of the study. There is no ranking within each set, items are ordered logically.
ARRIVE Essential 10		
Item 1 – Study design	Item 6 – Study design	We removed the reference to steps taken to minimise the effects of bias (formerly subitem 6b). All information about randomisation is now in item 4 and all information about blinding is now in item 5.
Item 2 – Sample size	Item 10 – Sample size	We clarified that the number of experimental units might be different from the number of animals. Independent replications are now mentioned with the results (item 10) to prevent any confusion with biological replicates.
Item 3 – Inclusion and exclusion criteria	Item 15 – Numbers analysed	We added a new subitem on a priori inclusion and exclusion criteria, evidence shows that ad hoc exclusion of data can lead to false positive results [1]. We clarified that the N number in each analysis might be different from the number of animals. We renamed the item to better reflect content.
Item 4 – Randomisation	Item 11 – Allocating animals to experimental groups	All references to randomisation were consolidated in this item for clarity. We reworded the text to include the randomisation procedure which was covered separately in the study design (formerly item 6). We clarified that experimental units are allocated to group, rather than animals.
Item 5 – Blinding	Item 6 – Study design	Blinding was included in the original guidelines as part of the study design (formerly subitem 6b), we have added more text in a new item to highlight its importance and encourage greater specificity.
Item 6 – Outcome measures	Item 12 – Experimental outcomes	We clarified that all outcome measures should be reported, and added a subitem to highlight the need to identify a primary outcome measure for hypothesis-testing studies. We changed the item name to 'outcome measures' because of concerns within the group that the term 'experimental outcomes' could be ambiguous.
Item 7 – Statistical methods	Item 13 – Statistical methods	We removed a reference to unit of analysis, which is often poorly understood and used the term experimental unit for consistency throughout the guidelines.
Item 8 – experimental animals	Item 8 – experimental animals	We clarified the wording and removed examples to streamline the guidelines, further details are discussed in the supporting E&E document [2].
Item 9 – Experimental procedures	Item 7 – Experimental procedures	We encouraged greater specificity by stating that procedures should be described in enough detail to allow others to replicate them. We removed examples to streamline the guidelines, further details are discussed in the supporting E&E document [2].
Item 10 – Results	Item 16 – outcomes and estimation	We expanded this item to provide more explicit guidance on reporting results. The name of the item was changed from 'outcomes and estimations' to 'results' for clarity and prevent confusion with item 6 – outcome measures.
Item removed	Item 14 – Baseline data	This item overlapped with item 8 – Experimental animals and the two items where combined, with further details provided in the supporting E&E document [2].

ARRIVE Recommended Set		
Item 11 – Abstract	Item 2 – Abstract	We specified that the sex of animals used should be included the abstract, empirical evidence suggests an endemic male bias in biomedical research [3].
Item 12 – Background	Item 3 – Background	We clarified the wording and removed examples to streamline the guidelines, further details are discussed in the supporting E&E document [2].
Item 13 – Objectives	Item 4 – Objectives	We removed a reference to primary and secondary objectives as it would not apply to exploratory studies, and added a requirement to describe the research question, which is relevant to all study types.
Item 14 – Ethical statement	Item 5 – Ethical statement	We removed reference to UK legislation to make this item relevant for an international audience. We added specification of the relevant licence or protocol numbers to provide accountability and promote transparency.
Item 15 – Housing and husbandry	Item 9 – Housing and husbandry	We moved the subitem on welfare-related assessments and interventions to item 16 – Animal care and monitoring. We removed examples to streamline the guidelines, further details are discussed in the supporting E&E document [2].
Item 16 – Animal care and monitoring	Item 17 – Adverse events	We added a new subitem to encourage the reporting of humane endpoints and monitoring, and changed the name of the item to animal care and monitoring to better reflect content.
Item 17 – Interpretation/scientific implications	Item 18 – Interpretation/scientific implications	We removed subitem c “Describe any implications of your experimental methods or findings for the replacement, refinement or reduction (3Rs) of the use of animals in research”. This item is not relevant to all animal studies and further details have been provided in the supporting E&E document [2].
Item 18 – Generalisability/ translation	Item 19 – Generalisability/ translation	We simplified and clarified the wording.
Item 19 – Protocol registration	New item	We added a new item on registering key aspects of the protocol. Empirical studies have shown up to 50% of outcomes which are measured are not reported [4]. This selective outcome reporting bias leads to an overstatement of biological effects.
Item 20 – Data access	New item	We added a new item on data access to encourage authors to provide a data sharing statement describing how others can gain access to the data on which the paper is based.
Item 21 – Declaration of interests	Item 20 – Funding	We added a new sub-item on declaring potential conflicts of interest. We added the specification of the role of the funder(s) in the ‘design, analysis and reporting of the study’. This information allows the reader to assess any competing interests, and any potential sources of bias. We renamed the item to better reflect content.
Item removed	Item 1 – Title	We removed this item as it provided no specific guidance on what to include in the title.

References

1. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22(11):1359-66. Epub 2011/10/19. doi: 10.1177/0956797611417632. PubMed PMID: 22006061.

2. Percie du Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, et al. Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2019. Submitted to Plos Biology. 2019.

3. Beery AK, Zucker I. Sex bias in neuroscience and biomedical research. Neurosci Biobehav Rev. 2011;35(3):565-72. Epub 2010/07/14. doi: 10.1016/j.neubiorev.2010.07.002. PubMed PMID: 20620164; PubMed Central PMCID: PMCPMC3008499.

4. Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, et al. Evaluation of excess significance bias in animal studies of neurological diseases. PLoS Biol. 2013;11(7):e1001609. Epub 2013/07/23. doi: 10.1371/journal.pbio.1001609. PubMed PMID: 23874156; PubMed Central PMCID: PMC3712913.

S2 – Delphi methods and results

1 Methods

1.1 Development of the revised checklist

For each subitem of ARRIVE 2010, the NC3Rs summarised the evidence justifying its inclusion in the guidelines and any indication of a need for revision. The working group then met for a two-day meeting in November 2017 in London, to review this information, discuss the addition of new items and agree on the strategy to go forward. We agreed to update the guidelines, develop an explanation and elaboration (E&E) document for ARRIVE 2019 [1] and prioritise the items to facilitate the uptake of the revised guidelines [2]. After the meeting, each item was allocated to at least two members of the working group to develop the item's explanation in more detail and refine the item's wording. Further iterations of the checklist were achieved by email discussion within the whole group.

The Delphi exercise was designed to achieve consensus on prioritising items of the ARRIVE guidelines.

The objective was to allocate the 22 items into two or three shortlists with different levels of priority, and relatively even distribution within each set.

1.2 Recruitment of the Delphi expert panel

Ethical approval for this study was obtained from the University of Bristol, Faculty of Science Research Ethics Committee (ID 66625).

The panel consisted of the ARRIVE Working Group and external stakeholders nominated by the Working Group, with suitable expertise on the quality of animal research or its reporting. We aimed to gather a diverse panel of experts, both in terms of field of expertise and geographical location.

Panel members consented to take part by following a link in the invitation email to the first round.

1.3 The Delphi process

There were three iterations of the questionnaire in total [3], and these were managed using the Comet Initiative DelphiManager platform (http://www.comet-initiative.org/DelphiManager/). Data collection took place June to November 2018. Panel members received an email invitation at the start of each round with a link to the online questionnaire. They were allowed three weeks to complete the questionnaire, with email reminders at day 7 and day 14. If they did not respond within the time frame they were excluded from that round, however they were invited to take part in the subsequent rounds of the Delphi.

Each of the 22 items of the revised ARRIVE guidelines was evaluated against the statement:

"How important is this piece of information for assessing the reliability of results in an animal research paper?"

Panel members scored each item on a scale of 1 – 9, where 1 was least important and 9 was most important.

The questionnaire presented in round 1 included free-text fields to provide reasoning for the score given to each item. Individual justifications were collated, summarised and presented to the whole panel in round 2.

In round 2, panel members were asked to provide a justification if their score for a particular item had changed between round 1 and round 2. Similarly, this information was summarised and presented to the whole panel in round 3.

Following rounds 1 and 2, the scores for each item were analysed and a structured summary consisting of a histogram showing the dispersal of the scores in the entire panel was prepared. This summary was presented with a new iteration of the questionnaire at the next round, where panel members were asked to re-score the items. In round 2 and 3, panel members' own scores from the previous round were also displayed for each item.

To encourage a wider dispersal of scores, in the final round (round 3) panel members were asked to follow two rules while scoring items:

- to score no more than ten items in the top range (7 – 9)
- to score no fewer than six items in the bottom range (1 – 3)

In the final dataset we excluded data entries which had not followed these rules, allowing for a deviation of ±1 item in each range.

1.4 Addition of new ARRIVE items

In the first round of the Delphi, we asked the panel to suggest new items that they believed should be included in the revised guidelines. The threshold for inclusion was defined a priori; for a new item to be considered, it would have to be suggested by at least 10% of the panel. The panel also had the opportunity to provide general feedback at the end of the survey.

1.5 Criteria for allocating items to sets

The plan to achieve consensus was defined a priori and two options were considered. The first option was to allocate the items in three sets, based on each item's median score and a minimum of 70% of the panel scoring the item within the same range. Score ranges were defined as follows:

- top range (7 – 9)
- middle range (4 – 6)
- bottom range (1 – 3)

Should the panel fail to reach agreement using the first option, the second option was to allocate the items in two sets and allocate items with a median score of 7 or above and an agreement level greater than 70% to the first set, and all other items to the second set.

Once data collection was completed, the ARRIVE working group met via videoconference to review the results and discuss the allocation of items into sets. As only 10 of 22 items reached the predefined agreement consensus of 70% (see supplementary information S3 – Delphi results), the second option was used to allocate items into two sets.

2 Results

2.1 Composition of the Delphi expert panel

One hundred experts were invited to participate in the Delphi exercise, 73 accepted the invitation and 71 participated in the final round (see Figure 1). Ten data entries, which did not follow data dispersal rules were excluded, 61 data entries were therefore included in the final score analysis.

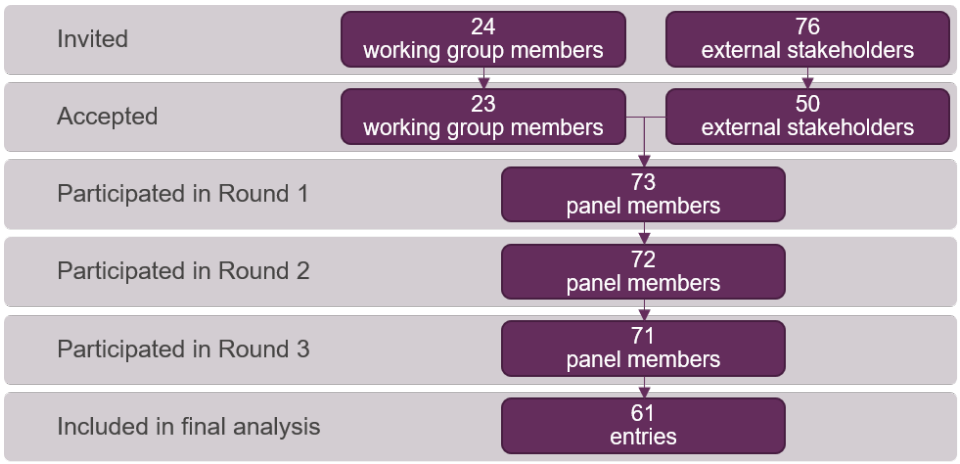


Figure 1. Delphi panel included at each stage of the Delphi exercise.

Demographics of the panel are presented in Table 2.

2.2 Suggestions for new items and feedback on existing items

18 panel members suggested a total of 31 new items (see supplementary information S3 – Delphi data). No new item suggestion met the 10% threshold for inclusion in the revised guidelines.

Feedback on the wording of existing items was considered by the working group in the drafting of the revised items and the drafting of the accompanying E&E document.

Feedback from the Delphi panel indicated that the item on number analysed was misunderstood and confused with the item on sample size. For clarity, the item on number analysed was incorporated to the item on inclusion and exclusion criteria in further iterations of the guidelines. This reduced the number of items to 21.

2.3 Scores for each Delphi round

The scores assigned to each item in rounds 1, 2 and 3 are shown in Figure 2.

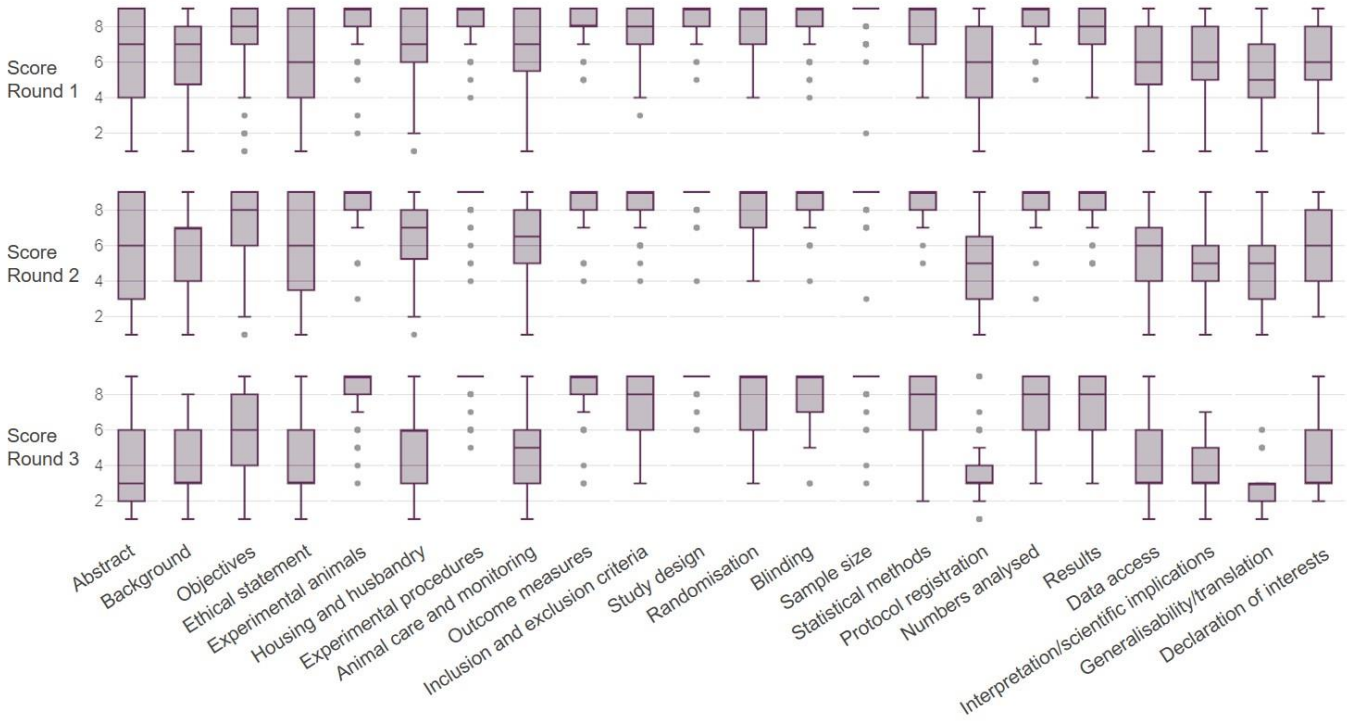


Figure 2. Item scores for each of the three Delphi rounds. Box and whisker plots of the panel members' scores for the 22 items. Round 1: n=71-73, round 2: n=70-71, round 3: n=61, the exact sample size for each item in each round is provided in supplementary information S3 – Delphi data. Data plotted as median, interquartile range, minimum, maximum and outliers using https://www.displayr.com/. Raw data available at https://osf.io/8xjdr/.

2.4. Allocation of items to sets

The allocation of items into sets is presented in Table 1. Eight items were shortlisted based on the a priori criterion (score in the top range and over 70% agreement within the panel). Three further items scoring in the top range were added to the shortlist following discussion within the working group.

Note that 11 items were allocated to set 1 but the combination of inclusion and exclusion criteria and numbers analysed in subsequent iterations of the guidelines reduced that number to 10 shortlisted items.

Set	Item	Score	% scores in the top range	Reasoning
Set 1	Study design	9 (9-9)	95	All items met pre-defined threshold (70%) for Set 1.
	Sample size	9 (9-9)	92	
	Experimental procedures	9 (9-9)	87	
	Outcome measures	9 (8-9)	85	
	Experimental animals	9 (8-9)	84	
	Blinding	9 (7-9)	77	
	Randomisation	9 (6-9)	70	
	Statistical methods	8 (6-9)	72	
	Numbers analysed	8 (6-9)	69	Median score in the top range.
	Inclusion and exclusion criteria	8 (6-9)	66	
	Results	8 (6-9)	59	
Set 2	Objectives	6 (4-8)	33	Median score outside the top range.
	Housing and husbandry	6 (3-6)	21	
	Animal care and monitoring	5 (3-6)	10	
	Ethical statement	3 (3-6)	20	
	Abstract	3 (2-6)	16	
	Data Access	3 (3-6)	11	
	Background	3 (3-6)	8	
	Protocol registration	3 (3-4)	7	
	Interpretation/Scientific implications	3 (3-5)	2	
	Declaration of interests	3 (3-6)	2	
	Generalisability/Translation	3 (2-3)	0	

Table 1. Allocation of the 22 items into two sets. Scores are displayed as median and interquartile range (IQR), n=61.

Primary country of work	
UK	27
USA	11
Canada	7
Brazil	5
Australia	3
China	2
Germany	2
Hong Kong	2
Switzerland	2
The Netherlands	2
Argentina	1
Belgium	1
India	1
Japan	1
Korea	1
Nigeria	1
Norway	1
South Africa	1
South Korea	1
Sri Lanka	1

Sector of work	
Academia	45
Not-for-profit	7
Industry	6
Publishing	6
Government	5
Funding body	1
Media	1
Educator	1
Contract research company	1

Years of relevant experience	
1 - 10	13
11 - 20	27
21 - 30	22
31 - 40	8
41 - 50	3

Career level	
Principle Investigator	32
Senior Staff	5
Associate Professor	4
Director	4
Staff Scientist	3
Editor	2
Executive Editor	2
Postdoctoral scientist	2
Senior Editor	2
Associate Director	1
Associate Editorial Director	1
Editorial Director	1
Head of Department	1
Head of Laboratory and Research	1
Head of Policy	1
Lab Animal Facility Director	1
Manager of Clinical Phenotyping Core	1
Managing Editor	1
Masters/PhD student	1
Mid-management	1
Policy analyst	1
Science Administrator	1
Senior Manager	1
No answer	3

Professional role	60
In vivo researcher	32
Journal editor	20
Statistician	6
Professor	5
Systematic review/ meta-researcher	4
Veterinarian/ assistant veterinarian	4
Director of a lab animal facility	3
Former *in vivo* researcher	3
Educator	2
Project manager	2
Publisher	2
Reviewer of in vivo research	2
Associate editorial director	1
Clinician	1
Director of research policy for a research funder	1
Director of research quality for a research funder	1
Director of standards	1
Evidence synthesis specialist	1
Head of experimental design for a research funder	1
Mathematical biologist	1
Op-ed editor	1
Policy analyst for a research funder	1
Preclinical bioresearch quality & compliance	1
Researcher using in vitro methods and human subjects	1
Science administrator	1
Scientific director	1
Scientist/manager	1
Secretary of a 3Rs centre	1
Senior program officer (science policy)	1

Table 2. Demographics of Delphi respondents (n = 73). Note that the total number of professional roles exceeds the number of panel members as they could select more than one role.

References

1. Percie du Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, et al. Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2019. Submitted to Plos Biology. 2019.

2. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Altman DG, Avey MT, et al. Revision of the ARRIVE guidelines: Rationale and scope. BMJ Open Science. 2018;2(1).

3. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the delphi method for selecting healthcare quality indicators: A systematic review. PLoS One. 2011;6(6):e20476. Epub 2011/06/23. doi: 10.1371/journal.pone.0020476. PubMed PMID: 21694759; PubMed Central PMCID: PMCPMC3111406.

Reporting animal research: Explanation and Elaboration for the ARRIVE guidelines 2019

Nathalie Percie du Sert[1], Amrita Ahluwalia[2], Sabina Alam[3], Marc T. Avey[4], Monya Baker[5], William J. Browne[6], Alejandra Clark[7], Innes C. Cuthill[6], Ulrich Dirnagl[8], Michael Emerson[9], Paul Garner[10], Stephen T. Holgate[11], David W. Howells[12], Viki Hurst[1], Natasha A. Karp[13], Katie Lidster[1], Catriona J. MacCallum[14], Malcolm Macleod[15], Esther J Pearl[1], Ole Petersen[16], Frances Rawle[17], Penny Reynolds[18], Kieron Rooney[19], Emily S. Sena[15], Shai D. Silberberg[20], Thomas Steckler[21], Hanno Würbel[22]

1. NC3Rs, UK
2. Queen Mary University of London, UK
3. Taylor & Francis Group, UK
4. ICF, USA
5. Nature, USA
6. University of Bristol, UK
7. PLOS ONE, UK
8. Universitätsmedizin Berlin, Germany
9. Imperial College London, UK
10. Liverpool School of Tropical Medicine, UK
11. University of Southampton, UK
12. University of Tasmania, Australia
13. Data Sciences & Quantitative Biology, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK
14. Hindawi Ltd, UK
15. Centre for Clinical Brain Sciences, University of Edinburgh, UK
16. Cardiff University, UK
17. Medical Research Council, UK
18. University of Florida, USA
19. University of Sydney, Australia
20. National Institute of Neurological Disorders and Stroke, USA
21. Janssen Pharmaceutica NV, Belgium
22. University of Bern, Switzerland

Abstract

Improving the reproducibility of biomedical research is a major challenge. Transparent and accurate reporting are vital to this process; it allows readers to assess the reliability of the findings, and repeat or build upon the work of other researchers. The NC3Rs developed the ARRIVE guidelines in 2010 to help authors and journals identify the minimum information necessary to report in publications describing *in vivo* experiments.

Despite widespread endorsement by the scientific community, the impact of the ARRIVE guidelines on the transparency of reporting in animal research publications has been limited. We have revised the ARRIVE guidelines to update them and facilitate their use in practice. The revised guidelines are published alongside this paper. This Explanation and Elaboration document was developed as part of the revision. It provides further information about each of the 21 items in ARRIVE 2019, including the rationale and supporting evidence for their inclusion in the guidelines, elaboration of details to report, and examples of good reporting from the published literature.

Introduction

Transparent and accurate reporting is essential to improve the reproducibility of scientific research; it enables others to scrutinise the methodological rigour of the studies, assess how reliable the findings are, and repeat or build upon the work.

However, evidence shows that the majority of publications fail to include key information and there is significant scope to improve the reporting of studies involving animal research [1-4]. To that end, the NC3Rs published the ARRIVE guidelines in 2010. The guidelines are a checklist of information to include in a manuscript to ensure that publications contain enough information to add to the knowledge base [5]. The guidelines have received widespread endorsement from the scientific community and are currently recommended by more than a thousand journals, with further endorsement from research funders, universities and learned societies worldwide.

Studies measuring the impact of ARRIVE on the quality of reporting have produced mixed results [6-11] and there is evidence that *in vivo* scientists are not sufficiently aware of the importance of reporting the information covered in the guidelines, and fail to appreciate the relevance to their work or their research field [12].

As a new international working group – the authors of this publication, we have revised the guidelines to update them and facilitate their uptake; the ARRIVE guidelines 2019 are published alongside this paper [13]. We have updated the recommendations in line with current best practice, reorganised the information and classified the items into two sets. The ARRIVE Essential 10 constitute the minimum reporting requirement and the Recommended Set provides further context to the study described. The two sets help authors, journal staff, editors and reviewers use the guidelines in practice, and allow a pragmatic implementation with an initial focus on the most critical issues. Once the Essential 10 are consistently reported in manuscripts, items from the Recommended Set can be added to journal requirements over time until all 21 items are routinely reported in all manuscripts. Full methodology for the revision and the allocation of items into sets is described in the accompanying publication [13].

A key aspect of the revision was to develop this Explanation and Elaboration document to provide background and rationale for each of the 21 items of ARRIVE 2019. Here we present additional guidance for each item and subitem, explain the importance of reporting this information in manuscripts that describe animal research, elaborate on what to report, and provide supporting evidence. Each subitem is also illustrated with examples of good reporting from the published literature.

Box 1: Glossary

Bias: Introduction of a systematic error in the estimated effect of an intervention, caused by inadequacies in the design, conduct, or analysis of an experiment.

Effect size: Quantitative measure that estimates the magnitude of differences between groups, or relationships between variables.

Experimental unit: Biological entity subjected to an intervention independently of all other units, such that it is possible to assign any two experimental units to different treatment groups.

External validity: Extent to which the results of an animal experiment provide a correct basis for generalisations to other populations of animals (including humans) and/or other environmental conditions.

False positive: Statistically significant result obtained by chance when the effect being investigated does not exist.

False negative: Non-statistically significant result obtained when the effect being investigated genuinely exists.

Independent variable of interest: Factor that a researcher manipulates within a controlled environment in order to test its impact on the outcome measured. Also known as: predictor variable, factor of interest.

Internal validity: Refers to the rigour of the study design and statistical analysis to isolate cause and effect, and attribute the effect observed to manipulation of the independent variable of interest. In an experiment with high internal validity, sources of bias and chance observations are minimised. In an experiment with low internal validity, the effect may be caused by bias, chance and other nuisance variables rather than the independent variable(s) of interest.

Null and alternative hypotheses: The null hypothesis (H_0) refers to the postulate that the response being measured is unaffected by the experimental manipulation being tested. The alternative hypothesis (H_1) refers to the postulate that manipulating the independent variable of interest has an effect on the response measured.

Nuisance variable: Sources of variability or conditions that could potentially bias results. Also known as: confounding factor, confounding variable

Outcome measure: Any variable recorded during a study to assess the effects of a treatment or experimental intervention. Also known as: dependent variable, response variable

Power: Probability that a test of significance will detect an effect (i.e. a deviation from the null hypothesis), if an effect exists (i.e. true positive result).

Sample size: Number of experimental units per group, also referred to as N number.

Definitions adapted from [14, 15] and placed in the context of animal research.

1. ARRIVE Essential 10

The ARRIVE Essential 10 (Box 2) constitute the minimum reporting requirement, to ensure that reviewers and readers can assess the reliability of the findings presented. There is no ranking within the set, items are presented in a logical order.

Box 2: ARRIVE Essential 10

1.	Study design
2.	Sample size
3.	Inclusion and exclusion criteria
4.	Randomisation
5.	Blinding
6.	Outcome measures
7.	Statistical methods
8.	Experimental animals
9.	Experimental procedures
10.	Results

Item 1. Study design

For each experiment, provide brief details of study design including:

1a. The groups being compared, including control groups. If no control group has been used, the rationale should be stated.

Explanation

The choice of control or comparator group is dependent on the experimental objective. Negative controls are used to determine if a difference between groups is caused by the intervention (e.g. wild-type animals vs genetically modified animals, placebo vs active treatment, sham surgery vs surgical intervention). Positive controls can be used to support the interpretation of negative results or determine if an expected effect is detectable.

It may not be necessary to include a separate control with no active treatment if, for example, the experiment aims to compare a treatment administered by different methods (e.g. intraperitoneal administration vs. oral gavage), or animals that are used as their own control in a longitudinal study. A pilot study, such as one designed to test the feasibility of a procedure might also not require a control group.

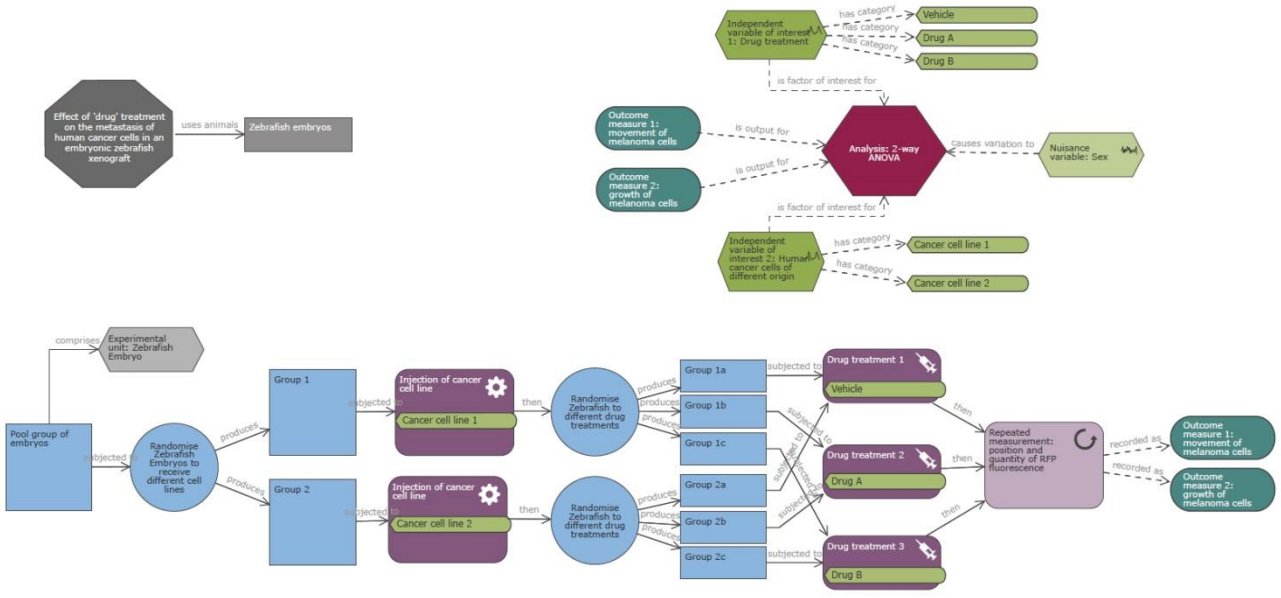
For complex study designs, a visual representation is more easily interpreted than a text description, so a timeline diagram or flow chart is recommended. Diagrams facilitate the identification of which treatments and procedures were applied to specific animals or groups of animals, and at what point in the study these were performed. They also help to communicate complex design features such as clustering or nesting (hierarchical designs), blocking (to reduce unwanted variation, see **item 4 – Randomisation**), or repeated measurements over time on the same experimental unit (repeated measures designs) [16, 17]. The Experimental Design Assistant (EDA) is a platform to support researchers in the design of in vivo experiments, it can be used to generate diagrams to represent any type of experimental design [18].

Report the groups clearly so that test groups, comparators and controls (negative or positive) can be identified easily. State clearly if the same control group was used for multiple experiments.

Examples

1) "The DAV1 study is a one-way, two-period crossover trial with 16 piglets receiving amoxicillin and placebo at period 1 and only amoxicillin at period 2. Amoxicillin was administered orally with a single dose of 30 mg.kg^{-1}. Plasma amoxicillin concentrations were collected at same sampling times at each period: 0.5, 1, 1.5, 2, 4, 6, 8, 10 and 12 h." [19]

2) "Figure A: Example of a study plan created using the Experimental Design Assistant showing a simple comparative study for the effect of two drugs on the metastatic spread of two different cancer cell lines. Block randomisation has been used to create 3 groups containing an equal number of zebrafish embryos injected with either cell line, and each group will be treated with a different drug treatment (including vehicle control). Each measurement outcome will be analysed by 2-way ANOVA to determine the effect of drug treatment on growth, survival and invasion of each cancer cell line." [20]



1b. The experimental unit (e.g. a single animal, litter, or cage of animals).

Explanation

The experimental unit is the biological entity subjected to an intervention independently of all other units, such that it is possible to assign any two experimental units to different treatment groups. The sample size is the number of experimental units per group.

Clearly indicate the experimental unit for each experiment so that the sample sizes and statistical analyses can be properly evaluated. There is a risk that if the experimental unit is not correctly identified, the sample size used in the data analysis will be incorrect. Inflation of the sample size by conflating experimental units with subsamples or repeated measurements is known as 'pseudoreplication'. This may invalidate the analysis and resulting conclusions [21, 22] (see also **item 7 – Statistical methods**).

Commonly, the experimental unit is the individual animal, each independently allocated to a treatment group (e.g. a drug administered by injection). However, the experimental unit may be the cage or the litter (e.g. a diet administered to a whole cage, or a treatment administered to a dam and investigated in her pups), or it could be part of the animal (e.g. different drug treatments applied topically to distinct body regions of the same animal). Animals may also serve as their own controls receiving different treatments separated by washout periods; here the experimental unit is an animal for a period of time. There may also be multiple experimental units in a single experiment, such as when a treatment such as diet is given to a pregnant dam and then the weaned pups are allocated to different diets [23]. See [24-26] for further guidance on identifying experimental units.

Examples

1) "The present study used the tissues collected at E15.5 from dams fed the 1X choline and 4X choline diets (n = 3 dams per group, per fetal sex; total n = 12 dams). To ensure statistical independence, only one placenta (either male or female) from each dam was used for each experiment. Each placenta, therefore, was considered to be an experimental unit." [27]
2) "We have used data collected from high-throughput phenotyping, which is based on a pipeline concept where a mouse is characterized by a series of standardized and validated tests underpinned by standard operating procedures (SOPs)…. The individual mouse was considered the experimental unit within the studies." [28]
3) "Fish were divided in two groups according to weight (0.7-1.2 g and 1.3-1.7 g) and randomly stocked (at a density of 15 fish per experimental unit) in 24 plastic tanks holding 60 L of water." [29]
4) "In the study, n refers to number of animals, with five acquisitions from each [corticostriatal] slice, with a maximum of three slices obtained from each experimental animal used for each protocol (six animals each group)." [30]

Item 2. Sample size

2a. Specify the exact number of experimental units allocated to each group, and the total number in each experiment. Also indicate the total number of animals used.

Explanation

Sample size relates to the number of experimental units in each group at the start of the study, and is usually represented by N (See **item 1 - Study design** for further guidance on identifying and reporting experimental units). This information is crucial to assess the validity of the statistical model and the robustness of the experimental results.

Report the exact value of N per group and the total number in each experiment. If the experimental unit is not the animal, also report the total number of animals to help readers understand the study design. For example, in a study investigating diet using cages of animals housed in pairs, the number of animals is double the number of experimental units. Reporting the total number of animals is also useful to identify if any were re-used between experiments.

Example

1) "Dams (for n see Table …) were assigned to treatments in a manner that provided similar means and variances in body weight before dosing was initiated…. For statistical purposes, the numbers/group are the number of litters, not the number of pups." [31]

Maternal and Neonatal Pup Effects of Various Phthalate Esters (PE) after Perinatal Maternal Exposure (GD 14–PND 3) to 0.75 g PE/kg/day on the Sprague-Dawley Rat

	Control	***BBP***	***DEHP***	***DINP***	***DEP***	***DMP***	***DOTP***
Numbers of dams assigned to this study Block 1/Block 2	*9/10*	*5/8*	*7/9*	*6/8*	*5/0*	*5/0*	*0/8*

Numbers of dams died	*0*	*0*	*0*	*0*	*2*	*1*	*0*	
Numbers of dams with live pups at day 2	*19*	*13*	*16*	*14*	*3*	*4*	*8*	
Numbers of dams with live pups at weaning	*19*	*11*	*16*	*14*	*3*	*4*	*8*	*[31]*

2b. Explain how the sample size was decided. Provide details of any *a priori* sample size calculation, if done.

Explanation

For any type of experiment, it is crucial to explain how the sample size was determined. For hypothesis-testing experiments, where inferential statistics are used to estimate the size of the effect and to determine the weight of evidence against the null hypothesis, the sample size needs to be justified to ensure experiments are of an optimal size to test the research question [32, 33] (see **item 13 – Objectives**). Power is the probability that a test of significance will detect an effect (i.e. a deviation from the null hypothesis), when the effect being investigated genuinely exists (i.e. true positive result). Sample sizes that are too small (i.e. underpowered studies) produce inconclusive results, whereas sample sizes that are too large (i.e. overpowered studies) raise ethical issues over unnecessary use of animals and may produce trivial findings that are statistically significant but not biologically relevant [34]. Low power has three effects: first, within the experiment, real effects are more likely to be missed; second, where an effect is detected, this will often be an over-estimation of the true effect size [25]; and finally, when low power is combined with publication bias, there is an increase in the false positive rate in the published literature [35]. Consequently, low powered studies contribute to the poor internal validity of research and risk wasting animals used in inconclusive research [36].

Study design can influence the statistical power of an experiment. Split-plot designs [37], factorial designs [38], or group-sequential designs [39] can increase the power of a study for a given number of animals. Statistical programs to help perform *a priori* sample size calculations exist for a variety of experimental designs and statistical analyses, for example G*power [40]. Choosing the appropriate calculator or algorithm to use depends on the type of outcome measures and independent variables, and the number of groups. Consultation with a statistician is recommended, especially when the experimental design is complex or unusual.

Where the experiment tests the effect of an intervention on the mean of a continuous outcome measure, the sample size can be calculated *a priori,* based on a mathematical relationship between the desired effect size, variability estimated from prior data, chosen significance level, power and sample size (See Box 3, and [24, 41] for practical advice). For an *a priori* sample size determination, report the analysis method (e.g. two-tailed student's t-test with a 0.05 significance threshold), the effect size of interest and a justification explaining why this effect size is relevant, the estimate of variability used (e.g. standard deviation) and how it was estimated, and the power selected.

Box 3: Information used in a power calculation

Sample size calculation is based on a mathematical relationship between the following parameters: effect size, variability, significance level, power and sample size. Questions to consider are:

The primary objective of the experiment – what is the main outcome measure?

The primary outcome measure should be identified in the planning stage of the experiment; it is the outcome of greatest importance, which will answer the main experimental question.

What is a biologically or clinically relevant effect size?

The effect size is the minimum change in the primary outcome measure between the groups under study, which would be of interest biologically and would be worth taking forward into further work.

What is the estimate of variability?

Estimates of variability can be obtained:

- From data collected from a preliminary experiment conducted under identical conditions to the planned experiment, e.g. a previous experiment in the same lab, testing the same treatment, under similar conditions, on animals with the same characteristics.
- From the control group in a previous experiment testing a different treatment.
- From a similar experiment reported in the literature.

What risk of a false positive is acceptable? (significance threshold)

The significance level or threshold (α) is the probability of obtaining a significant result by chance (a false positive) when the null hypothesis is true (i.e. there is no real, biologically relevant difference between the groups). If it is set at 0.05 then the risk of obtaining a false positive is 1 in 20 for a single statistical test. However, the threshold or the p values will need to be adjusted in scenarios of multiple testing (e.g. by using a Bonferroni correction).

What risk of a false negative is acceptable? (power)

The power (1-β) is the probability that the experiment will correctly lead to the rejection of the null hypothesis if the effect being investigated genuinely exists (i.e. detect that there is a biologically meaningful difference when there is one). A target power between 80-95% is normally deemed acceptable.

Will you use a one or two-sided test? (directionality)

The directionality of a test depends on the distribution of the test statistics for a given analysis. For tests based on t or z distributions (such as t-tests), whether the data will be analysed using a one or two-sided test relates to whether the alternative hypothesis (H_1) is directional or not. An experiment with a directional (one-sided) H_1 can be powered and analysed with a one-sided test. This assumes that direction of the effect is known (this is very rare in biology) and the goal is to maximise the chances of detecting this effect. However, the investigator cannot then test for the possibility of missing an effect in the untested direction. Choosing a one-tailed test for the sole purpose of attaining statistical significance is not appropriate.

Two-sided tests with a non-directional H_1 are much more common and allow researchers to detect the effect of a treatment regardless of its direction.

Note that analyses such as ANOVA and chi-square are based on asymmetrical distributions (F- distribution and chi-square distribution) with only one tail. Therefore, these tests do not have a directionality option.

There are several types of studies where *a priori* sample size calculations are not appropriate. For example, the number of animals needed for antibody or tissue production is determined by the amount required and the production ability of an individual animal. For studies where the outcome is a successful generation of a sample or a condition (e.g. the production of transgenic animals), the number of animals is determined by the probability of success of the experimental procedure.

In early feasibility or pilot studies, the number of animals required depends on the purpose of the study. Where the objective of the preliminary study is to improve procedures and equipment, the number of animals needed is generally small. In such cases power calculations are not appropriate and sample sizes can be estimated based on operational capacity and constraints [42]. Pilot studies alone are unlikely to provide adequate data on variability for a power calculation for future experiments. Systematic reviews and previous studies are more appropriate sources of information on variability [43].

Regardless of whether a power calculation was used or not, when explaining how the sample size was determined take into consideration any anticipated loss of animals or data, for example due to exclusion criteria established upfront or expected attrition (see **item 3 – inclusion and exclusion criteria**).

Examples

1) "The sample size calculation was based on postoperative pain numerical rating scale (NRS) scores after administration of buprenorphine (NRS AUC mean = 2.70; noninferiority limit = 0.54; standard deviation = 0.66) as the reference treatment and also Glasgow Composite Pain Scale (GCPS) scores using online software (Experimental design assistant; https://eda.nc3rs.org.uk/eda/login/auth). The power of the experiment was set to 80%. A total of 20 dogs per group were considered necessary." [44]

2) "We selected a small sample size because the bioglass prototype was evaluated in vivo for the first time in the present study, and therefore, the initial intention was to gather basic evidence regarding the use of this biomaterial in more complex experimental designs." [45]

Item 3. Inclusion and exclusion criteria

3a. Describe any criteria established *a priori* for including or excluding animals (or experimental units) during the experiment, and data points during the analysis.

Explanation

Inclusion and exclusion criteria define the eligibility or disqualification of animals and data once the study has commenced. To ensure scientific rigour, the criteria should be defined before the experiment starts and data are collected [46]. Inclusion criteria should not be confused with animal characteristics (see **item 8 – Experimental animals**) but can be related to these (e.g. body weights must be within a certain range for a particular procedure) or related to other study parameters (e.g. task performance has to exceed a given threshold). Exclusion criteria may result from technical or welfare issues such as complications anticipated during surgery, or circumstances where test procedures might be compromised (e.g. development of motor impairments that could affect behavioural measurements). Criteria for excluding samples or data include failure to meet quality control standards, such as insufficient sample volumes, unacceptable levels of contaminants, poor histological quality, etc. Similarly, how the researcher will define and handle data outliers during the analysis should be also decided before the experiment starts (see subitem 3b for guidance on responsible data cleaning).

Exclusion criteria may also reflect the ethical principles of a study in line with its humane endpoints (see **item 16 – Animal care and monitoring**). For example, in cancer studies an animal might be dropped from the study and euthanised before the predetermined time point if the size of a subcutaneous tumour exceeds a specific volume [47]. If losses are anticipated, these should be considered when determining the number of animals to include in the study (see **item 2 – Sample size**).

Best practice is to include all *a priori* inclusion and exclusion/outlier criteria in a pre-registered protocol (see **item 19 – Protocol registration**). At the very least these criteria should be documented in a lab notebook and reported in manuscripts, explicitly stating that the criteria were defined before any data was collected.

Example

1) "The animals were included in the study if they underwent successful MCA occlusion (MCAo), defined by a 60% or greater drop in cerebral blood flow seen with laser Doppler flowmetry. The animals were excluded if insertion of the thread resulted in perforation of the vessel wall (determined by the presence of sub-arachnoid blood at the time of sacrifice), if the silicon tip of the thread became dislodged during withdrawal, or if the animal died prematurely, preventing the collection of behavioral and histological data." [48]

3b. For each experimental group, report any animals, experimental units or data points not included in the analysis and explain why.

Explanation

Animals, experimental units, or data points that are unaccounted for can lead to instances where conclusions cannot be supported by the raw data [49]. Reporting exclusions and attritions provides valuable information to other investigators evaluating the results, or who intend to repeat the experiment or test the intervention in other species. It may also provide important safety information for human trials (e.g. exclusions related to adverse effects).

There are many legitimate reasons for experimental attrition, some of which are anticipated and controlled for in advance (see subitem 3a on defining exclusion and inclusion criteria) but some data loss might not be anticipated. For example, data points may be excluded from analyses due to an animal receiving the wrong treatment, unexpected drug toxicity, infections or diseases unrelated to the experiment, sampling errors (e.g. a malfunctioning assay that produced a spurious result, inadequate calibration of equipment), or other human error (e.g. forgetting to switch on equipment for a recording).

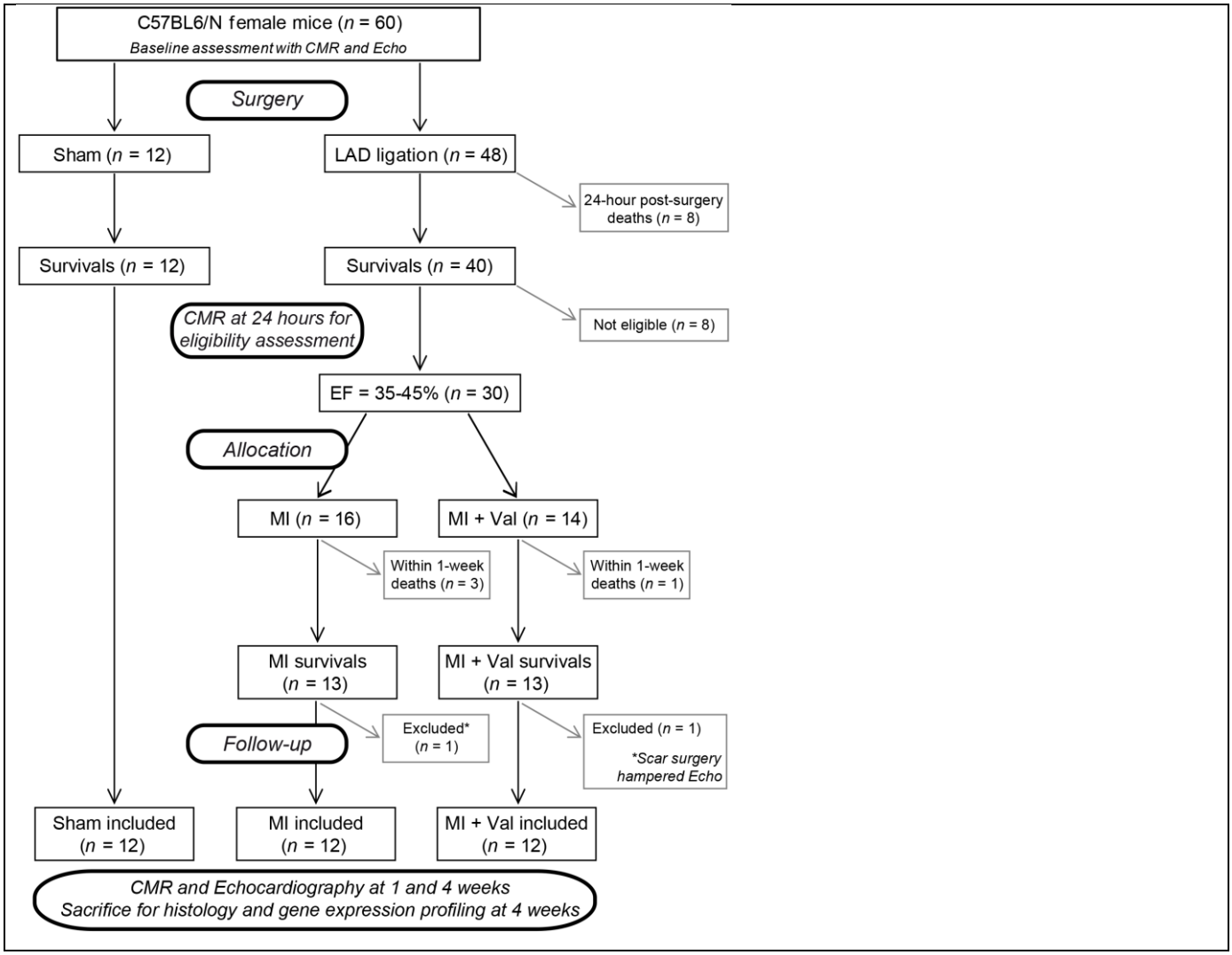
In some instances, it may be scientifically justifiable to remove outlier data points from an analysis, such as readings that are outside a plausible range. Providing the reasoning for removing data points enables the distinction to be made between responsible data cleaning and data manipulation. When reasons are not disclosed the reliability of the conclusions is in question, as inappropriate data cleaning has the potential to bias study outcomes [50].

There is a movement towards greater data sharing (see **item 20 – Data access**), along with an increase in strategies such as code sharing to enable analysis replication. These practices, however transparent, still need to be accompanied by a disclosure on the reasoning for data cleaning, and whether it was pre-defined.

Report all animal exclusions and loss of data points, along with the rationale for their exclusion. Accompanying these criteria should be an explicit description of whether researchers were blinded to the group allocations when data or animals were excluded (see **item 5 – Blinding** and [51]), and whether these criteria were decided prior to the experiment [8, 32, 52]. Explicitly state where built-in models in statistics packages have been used to remove outliers (e.g. GraphPad Prism's outlier test).

Examples

1) "Pen was the experimental unit for all data. One entire pen (ZnAA90) was removed as an outlier from both Pre-RAC and RAC periods for poor performance caused by illness unrelated to treatment.... Outliers were determined using Cook's D statistic and removed if Cook's D > 0.5. One steer was determined to be an outlier for day 48 liver biopsy TM and data were removed." [53]
2) "Seventy-two SHRs were randomized into the study, of which 13 did not meet our inclusion and exclusion criteria because the drop in cerebral blood flow at occlusion did not reach 60% (seven animals), postoperative death (one animal: autopsy unable to identify the cause of death), haemorrhage during thread insertion (one animal), and disconnection of the silicon tip of the thread during withdrawal, making the permanence of reperfusion uncertain (four animals). A total of 59 animals were therefore included in the analysis of infarct volume in this study. In error, three animals were sacrificed before their final assessment of neurobehavioral score: one from the normothermia/water group and two from the hypothermia/pethidine group. These errors occurred blinded to treatment group allocation. A total of 56 animals were therefore included in the analysis of neurobehavioral score." [48]
3) "Fig 1. Flow chart showing the experimental protocol with the number of animals used, died and included in the study…. After baseline CMR and echocardiography, MI was induced by left anterior descending (LAD) coronary artery ligation (n = 48), as previously described. As control of surgery procedure, sham operated mice underwent thoracotomy and pericardiotomy without coronary artery ligation (n = 12)."[54]



3c. For each analysis, report the exact value of N in each experimental group.

Explanation

The exact number of experimental units analysed in each group (i.e. the N number) is essential information for the reader to interpret the analysis, it should be reported unambiguously. All animals and data used in the experiment should be accounted for in the data presented. Sometimes, for good reasons, animals may need to be excluded from a study (e.g. illness or mortality), and data points excluded from analyses (e.g. biologically implausible values). Reporting losses will help the reader to understand the experimental design process, replicate methods, and provide adequate tracking of animal numbers in a study, especially when sample size numbers in the analyses do not match the original group numbers.

Indicate numbers clearly within the text or on figures, and provide absolute numbers (e.g. 10/20, not 50%). For studies where animals are measured at different time points, explicitly report the full description of which animals undergo measurement, and when [32].

Examples

1) "Group F contained 29 adult males and 58 adult females in 2010 (n = 87), and 32 adult males and 66 adult females in 2011 (n = 98). The increase in female numbers was due to maturation of juveniles to adults. Females belonged to three matrilines, and there were no major shifts in rank in the male hierarchy. Six mid to low ranking

individuals died and were excluded from analyses, as were five mid-ranking males who emigrated from the group at the beginning of 2011." [55]
2) "The proportion of test time that animals spent interacting with the handler (sniffed the gloved hand or tunnel, made paw contact, climbed on, or entered the handling tunnel) was measured from DVD recordings. This was then averaged across the two mice in each cage as they were tested together and their behaviour was not independent…. Mice handled with the home cage tunnel spent a much greater proportion of the test interacting with the handler (mean ± s.e.m., 39.8 ± 5.2 percent time of 60 s test, n = 8 cages) than those handled by tail (6.4 ± 2.0 percent time, n = 8 cages), while those handled by cupping showed intermediate levels of voluntary interaction (27.6 ± 7.1 percent time, n = 8 cages)." [56]

Item 4. Randomisation

Describe the methods used:

4a. To allocate experimental units to control and treatment groups. If randomisation was used, provide the method of randomisation.

Explanation

Using randomisation during the allocation to groups ensures that each experimental unit has an equal probability of receiving a particular treatment. It helps minimise selection bias and reduce systematic differences in the characteristics of animals allocated to different groups [57-59]. However, investigators frequently confuse "random" with "haphazard" or "arbitrary" allocation. Non-random allocation can introduce bias that influences the results, as a researcher may (consciously or subconsciously) make judgements in allocating an animal to a particular group, or because of unknown and uncontrolled differences in the experimental conditions or animals in different groups. Systematic reviews have shown that animal experiments that do not report randomisation or other bias-reducing measures such as blinding, are more likely to report exaggerated effects that meet conventional measures of statistical significance [60, 61]. It is especially important to use methods of randomisation in situations where it is not possible to blind all or parts of the experiment but even with randomisation, researcher bias can pervert the allocation. This can be avoided by using allocation concealment (**see item 5 – Blinding**). In studies where sample sizes are small, simple randomisation may result in unbalanced groups; here randomisation strategies to balance groups such as randomising in matched pairs [62-64] and blocking are encouraged [24].

Report the type of randomisation used (simple, stratified, randomised complete blocks, etc.; see Box 4), the method of randomisation (e.g. computer-generated randomisation sequence, with details of the algorithm or programme used), and what was randomised (e.g. treatment to experimental unit, order of treatment for each animal). The EDA has a dedicated feature for randomisation and allocation concealment [18].

Examples

1) "Fifty 12-week-old male Sprague-Dawley rats, weighing 320–360g, were obtained from Guangdong Medical Laboratory Animal Center (Guangzhou, China) and randomly divided into two groups (25 rats/group): the intact group and the castration group. Random numbers were generated using the standard = RAND() function in Microsoft Excel." [65]
2) "Animals were randomized after surviving the initial I/R, using a computer based random order generator." [66]
3) "At each institute, phenotyping data from both sexes is collected at regular intervals on age-matched wildtype mice of equivalent genetic backgrounds. Cohorts of at least seven homozygote mice of each sex per pipeline were generated…. The random allocation of mice to experimental group (wildtype versus knockout) was driven by Mendelian Inheritance." [28]

4b. To minimise potential confounding factors such as the order of treatments and measurements, or animal/cage location.

Explanation

Ensuring there is no systematic difference between animals in different groups apart from the experimental exposure is an important principle throughout the conduct of the experiment. Identifying nuisance variables (sources of variability or conditions that could potentially bias results), and managing them in the design of the experiment increases the sensitivity of the experiment. For example, rodents in cages at the top of the rack may be exposed to higher light levels, which can affect stress [67]. Mitigation strategies for nuisance variables include randomising or counterbalancing the position of animal cages on the rack, and taking measurements or processing samples in a random order (preferably with the investigator blinded to the treatment received; see **item 5 – Blinding**). Such practices help avoid introducing unintentional systematic differences between comparison groups, also known as order effects. Strategies to avoid order effects include counterbalancing, randomising order of treatment, and blocking (see Box 4).

Box 4: Considerations for randomisation

Simple randomisation

All animals/samples are simultaneously randomised to the treatment groups without considering any other variable. This strategy is rarely appropriate as it cannot ensure that comparison groups are balanced for factors or covariates that might influence the result of an experiment.

Randomisation within blocks

Blocking is a method of controlling natural variation among experimental units. This splits up the experiment into smaller sub-experiments (blocks), and treatments are randomised to experimental units within each block [24, 68, 69]. This takes into account nuisance variables that could potentially bias the results (e.g. cage location, day or week of procedure).

Stratified randomisation uses the same principle as randomisation within blocks, only the strata tend to be traits of the animal that are likely to be associated with the response (e.g. weight class or tumour size class). This can lead to differences in the practical implementation of stratified randomisation as compared to block randomisation (e.g. there may not be equal numbers of experimental units in each weight class).

Other randomisation techniques

Minimisation is an alternative strategy to allocate animals/samples to treatment group to balance confounding variables that might influence the result of an experiment. With minimisation the treatment allocated to the next animal/sample depends on the characteristics of those animals/samples already assigned. The aim is that each allocation should minimise the imbalance across multiple factors [70]. This approach works well for a continuous nuisance variable such as body weight or starting tumour volume.

Examples of nuisance variables that can be accounted for in the randomisation strategy

- Time or day of the experiment
- Litter, cage or fish tank
- Investigator or surgeon – different level of experience in the people administering the treatments, performing the surgeries, or assessing the results may result in varying stress levels in the animals or duration of anaesthesia
- Equipment (e.g. PCR machine, spectrophotometer) – calibration may vary
- Measurement of a study parameter (e.g. initial tumour volume)
- Animal characteristics – sex, age class, weight class
- Location – exposure to light, ventilation and disturbances may vary in cages located at different height or on different racks, which may affect important physiological processes

Implication for the analysis

If blocking factors are used in the randomisation, they should also be included in the analysis. Nuisance variables increase variability in the sample, which reduces statistical power. Including a nuisance variable as a blocking factor in the analysis accounts for that variability and can increase the power, thus increasing the ability to detect a real effect with fewer experimental units. However, blocking uses up degrees of freedom and thus reduces the power if the nuisance variable does not have a substantial impact on variability.

Report the methods used to minimise confounding factors alongside the methods used to allocate animals to groups.

Examples

1) "Randomisation was carried out as follows. On arrival from El-Nile Company, animals were assigned a group designation and weighed. A total number of 32 animals were divided into four different weight groups (eight animals per group). Each animal was assigned a temporary random number within the weight range group. On the basis of their position on the rack, cages were given a numerical designation. For each group, a cage was selected randomly from the pool of all cages. Two animals were removed from each weight range group and given their permanent numerical designation in the cages. Then, the cages were randomized within the exposure group." [71]
2) "...test time was between 08.30am to 12.30pm and testing order was randomized daily, with each animal tested at a different time each test day." [72]
3) "Bulls were blocked by BW into four blocks of 905 animals with similar BW and then within each block, bulls were randomly assigned to one of four experimental treatments in a completely randomized block design resulting in 905 animals per treatment. Animals were allocated to 20 pens (181 animals per pen and five pens per treatment)." [73]

Item 5. Blinding

Describe who was aware of the group allocation at the different stages of the experiment (during the allocation, the conduct of the experiment, the outcome assessment, and the data analysis).

Explanation

Researchers often expect a particular outcome, and can unintentionally influence the experiment or interpret the data in such a way as to support their preferred hypothesis [74]. Blinding is a strategy used to minimise these subjective biases.

Whilst there is primary evidence of the impact of blinding in the clinical literature that directly compares blinded vs unblinded assessment of outcomes [75], there is limited empirical evidence in animal research [76, 77]. There are, however, compelling data from systematic reviews showing that non-blinded outcome assessment leads to the treatment effects being overestimated, and the lack of bias-reducing measures such as randomisation and blinding can contribute to as much as 30-45% inflation of effect sizes [60, 78, 79].

Ideally, investigators should be unaware of the treatment(s) animals have received or will be receiving, from the start of the experiment until the data have been analysed. If this is not possible for all the stages of an experiment (see Box 5), it should always be possible to conduct at least some of the stages blind. This has implications for the organisation of the experiment and may require help from additional personnel, for example a surgeon to perform interventions, a technician to code the treatment syringes for each animal, or a colleague to code the treatment groups for the analysis. Online resources are available to facilitate allocation concealment and blinding [18].

Box 5: Blinding during different stages of an experiment

During allocation

Allocation concealment refers to concealing the treatment to be allocated to each individual animal from those assigning the animals to groups, until the time of assignment. Together with randomisation, allocation concealment helps minimise selection bias, which can introduce systematic differences between treatment groups.

During the conduct of the experiment

Where possible, animal care staff and those who administer treatments should be unaware of allocation groups to ensure that all animals in the experiment are handled, monitored and treated in the same way. Treating different groups differently based on the treatment they have received could alter animal behaviour and physiology, and produce confounds.

Welfare or safety reasons may prevent blinding of animal care staff but in most cases, blinding is possible. For example, if hazardous microorganisms are used, control animals can be considered as dangerous as infected animals. If a welfare issue would only be tolerated for a short time in treated but not control animals, a harm-benefit analysis is needed to decide whether blinding should be used.

During the outcome assessment

The person collecting experimental measurements or conducting assessments should not know which treatment each sample/animal received, and which samples/animals are grouped together. Blinding is especially important during outcome assessment, particularly if there is a subjective element (e.g. when assessing behavioural changes or reading histological slides) [76]. Randomising the order of examination can also reduce bias.

If the person assessing the outcome cannot be blinded to the group allocation (e.g. obvious phenotypic or behavioural differences between groups) some, but not all, of the sources of bias could be mitigated by sending data for analysis to a third party, who has no vested interest in the experiment and does not know whether a treatment is expected to improve or worsen the outcome.

During the data analysis

The person analysing the data should know which data are grouped together to enable group comparisons, but should not be aware of which treatment each group received. This type of blinding is often neglected, but is important as the analyst makes many semi-subjective decisions such as applying data transformation to outcome measures, choosing methods for handling missing data and handling outliers. How these decisions will be made should also be decided *a priori*.

Data can be coded prior to analysis so that the treatment group cannot be identified before analysis is completed.

Specify whether blinding was used or not for each step of the experimental process (see Box 5). If blinding was not possible during a specific stage of the experiment, provide the reason why.

Examples

1) "For each animal, four different investigators were involved as follows: a first investigator (RB) administered the treatment based on the randomization table. This investigator was the only person aware of the treatment group allocation. A second investigator (SC) was responsible for the anaesthetic procedure, whereas a third investigator (MS, PG, IT) performed the surgical procedure. Finally, a fourth investigator (MAD) (also unaware of treatment) assessed GCPS and NRS, MNT, and sedation NRS scores." [80]

2) "…due to overt behavioral seizure activity the experimenter could not be blinded to whether the animal was injected with pilocarpine or with saline." [81]

3) "Investigators could not be blinded to the mouse strain due to the difference in coat colors, but the three-chamber sociability test was performed with ANY-maze video tracking software (Stoelting, Wood Dale, IL, USA) using an overhead video camera system to automate behavioral testing and provide unbiased data analyses. The one-chamber social interaction test requires manual scoring and was analyzed by an individual with no knowledge of the questions." [82]

Item 6. Outcome measures

6a. Clearly define all outcome measures assessed (e.g. cell death, molecular markers, or behavioural changes).

Explanation

An outcome measure (also known as a dependent variable or a response variable) is any variable recorded during a study (e.g. volume of damaged tissue, number of dead cells, specific molecular marker) to assess the effects of a treatment or experimental intervention. Outcome measures may be important for characterising a sample (e.g. baseline data) or for describing complex responses (e.g. 'haemodynamic' outcome measures including heart rate, blood pressure, central venous pressure, and cardiac output).

Explicitly describe what was measured, especially when measures can be operationalised in different ways. For example, activity could be recorded as time spent moving or distance travelled. Where possible, the recording of outcome measures should be made in an unbiased manner (e.g. blinded to the treatment allocation of each

experimental group; see **item 5 – Blinding**). Specify how the outcome measure(s) assessed are relevant to the objectives of the study.

Example

1) "The following parameters were assessed: threshold pressure (TP; intravesical pressure immediately before micturition); post-void pressure (PVP; intravesical pressure immediately after micturition); peak pressure (PP; highest intravesical pressure during micturition); capacity (CP; volume of saline needed to induce the first micturition); compliance (CO; CP to TP ratio); frequency of voiding contractions (VC) and frequency of non-voiding contractions (NVCs)." [83]

6b. For hypothesis-testing studies, specify the primary outcome measure, i.e. the outcome measure that was used to determine the sample size.

Explanation

In a hypothesis-testing experiment, the primary outcome measure answers the main biological question. It is the outcome of greatest importance, identified in the planning stages of the experiment and used as the basis for the sample size calculation (see Box 3). For exploratory studies it is not necessary to identify a single primary outcome and often multiple outcomes are assessed (see **item 13 – Objectives**).

In a hypothesis-testing study powered to detect an effect on the primary outcome measure, data on secondary outcomes are used to evaluate additional effects of the intervention but subsequent statistical analysis of secondary outcome measures may be underpowered, making results and interpretation less reliable [84, 85]. Studies that claim to test a hypothesis but do not specify a pre-defined primary outcome measure, or those that change the primary outcome measure after data were collected (also known as primary outcome switching) are liable to selectively report only statistically significant results, favouring more positive findings [86].

Registering a protocol in advance protects the researcher against concerns about selective outcome reporting (also known as data dredging or p-hacking) and provides evidence that the primary outcome reported in the manuscript accurately reflects what was planned [87] (see **item 19 – Protocol registration**).

If the study was designed to test a hypothesis and more than one outcome was assessed, explicitly identify the primary outcome measure and state if it was defined as such prior to data collection.

Examples

1) "The primary outcome of this study will be forelimb function assessed with the staircase test. Secondary outcomes constitute Rotarod performance, stroke volume (quantified on MR imaging or brain sections, respectively), diffusion tensor imaging (DTI) connectome mapping, and histological analyses to measure neuronal and microglial densities, and phagocytic activity… The study is designed with 80% power to detect a relative 25% difference in pellet-reaching performance in the Staircase test." [88]

2. "The primary endpoint of this study was defined as left ventricular ejection fraction (EF) at the end of follow-up, measured by magnetic resonance imaging (MRI). Secondary endpoints were left ventricular end diastolic volume and left ventricular end systolic volume (EDV and ESV) measured by MRI, infarct size measured by ex vivo gross macroscopy after incubation with triphenyltetrazolium chloride (TTC) and late gadolinium enhancement (LGE) MRI, functional parameters serially measured by pressure volume (PV-)loop and echocardiography, coronary microvascular function by intracoronary pressure- and flow measurements and vascular density and fibrosis on histology. Based on a power calculation (estimated effect 7.5% [6], standard deviation of 5%, a power of 0.9 and alpha of 0.05) 8 pigs per group were needed." [66]

Item 7. Statistical methods

7a. Provide details of the statistical methods used for each analysis.

Explanation

In hypothesis-testing studies comparing two or more groups, inferential statistics are used to estimate the size of the effect and to determine the weight of evidence against the null hypothesis. The effect size is the magnitude of the difference between two groups. The description of the statistical analysis should provide enough detail so that another researcher could re-analyse the raw data using the same method and obtain the same results. Relevant information includes what the outcome measures and independent variables were, what statistical analyses were performed, what tests were used to check assumptions, and any data transformations [89]. Give details of any confounders, blocking factors or covariates taken into account for each statistical test, include how the effects of each were mitigated. This allows readers to assess if analysis methods were appropriate.

In exploratory studies where no specific hypothesis was tested, descriptive statistics can be used to summarise the data (see **item 10 – Results**). They do not allow conclusions beyond the data but are important for generating new hypotheses that may be tested in subsequent experiments.

For any study reporting descriptive statistics, explicitly state which measure of central tendency is reported (e.g. mean or median) and which measure of variability is reported (e.g. standard deviation, range, quartiles or interquartile range).

Examples

1) "Analysis of variance was performed using the GLM procedure of SAS (SAS Inst., Cary, NC). Average pen values were used as the experimental unit for the performance parameters. The model considered the effects of block and dietary treatment (5 diets). Data were adjusted by the covariant of initial body weight. Orthogonal contrasts were used to test the effects of SDPP processing (UV vs no UV) and dietary SDPP level (3% vs 6%). Results are presented as least squares means. The level of significance was set at P < 0.05." [90]
2) "All risk factors of interest were investigated in a single model. Logistic regression allows blocking factors and explicitly investigates the effect of each independent variable controlling for the effects of all others.... As we were interested in husbandry and environmental effects, we blocked the analysis by important biological variables (age; backstrain; inbreeding; sex; breeding status) to control for their effect. (The role of these biological variables in barbering behavior, particularly with reference to barbering as a model for the human disorder trichotillomania, is described elsewhere: Garner et al., 2004). We also blocked by room to control for the effect of unknown environmental variables associated with this design variable. We tested for the effect of the following husbandry and environmental risk factors: cage mate relationships (i.e. siblings, non-siblings, or mixed); cage type (i.e. plastic or steel); cage height from floor; cage horizontal position (whether the cage was on the side or the middle of a rack); stocking density; and the number of adults in the cage. Cage material by cage height from floor; and cage material by cage horizontal position interactions were examined, and then removed from the model as they were nonsignificant. N = 1959 mice were included in this analysis" [91]

7b. Specify the experimental unit that was used for each statistical test.

Explanation

Incorrect identification of the experimental unit can lead to pseudoreplication and underpowered studies (see **item 1 – Study design**). For example, measurements from 50 individual cells from a single mouse represent N = 1 when the experimental unit is the mouse. The 50 measurements are subsamples and provide an estimate of measurement error so should be averaged or used in a nested analysis. Reporting N = 50 in this case is an example of pseudoreplication [22]. It underestimates the true variability in a study, which can lead to false positives. If, however, each cell taken from the mouse is then randomly allocated to different treatments and assessed individually, the cell might be regarded as the experimental unit.

Explicitly report the experimental unit used in each statistical analysis.

Examples

1) "For each test, the experimental unit was an individual animal." [92]
2) "Maternal data regarding body weight, food intake, water consumption, urinary cotinine level and hormonal analysis were analyzed using the individual animal as the experimental unit. The data for offspring regarding body

weight, food intake, organ weight at necropsy, urinary cotinine level, immunohistochemical cellular distribution, TUNEL+ cells, RT-PCR and hormonal analysis were analyzed using the litter as the experimental unit." [93]

7c. Describe any methods used to assess whether the data met the assumptions of the statistical approach.

Explanation

Hypothesis tests are based on assumptions about the underlying data. Describing how assumptions were assessed, and whether these assumptions are met by the data, enables readers to assess the suitability of the statistical approach used. If the assumptions are incorrect, the conclusions may not be valid. For example, the assumptions for data used in parametric tests (such as a t-test, Z-test, ANOVA, Pearson's r coefficient, etc.) are that the data are continuous, the residuals from the analysis are normally distributed, the responses are independent, and that different groups should have similar variances.

There are various tests for normality, for example the Shapiro-Wilk and Kolmogorov-Smirnov tests. However, these tests have to be used cautiously. If the sample size is small, they will struggle to detect non-normality, if the sample size is large, the tests will detect minor deviations. An alternative approach is to evaluate data with visual plots e.g. normal probability plots, box plots, scatterplots. If the residuals of the analysis are not normally distributed, the assumption may be satisfied using a data transformation where the same mathematical function is applied to all data points to produce normally distributed data (e.g. $\log_e$, $\log_{10}$, square root, arcsine).

Other types of outcome measures (binary, categorical, or ordinal) will require different methods of analysis, and each will have different sets of assumptions. For example, categorical data are summarised by counts and percentages or proportions, and are analysed by tests of proportions; these analysis methods assume that data are binary, ordinal or nominal, and independent [94].

Report the type of outcome measure and the methods used to test the assumptions of the statistical approach. If data were transformed, identify precisely the transformation used and which outcome measures it was applied to.

Examples

1) "Model assumptions were checked using the Shapiro-Wilk normality test and Levene's Test for homogeneity of variance and by visual inspection of residual and fitted value plots. Some of the response variables had to be transformed by applying the natural logarithm or the second or third root, but were back-transformed for visualization of significant effects." [95]
2) The effects of housing (treatment) and day of euthanasia on cortisol levels were assessed by using fixed-effects 2-way ANOVA. An initial exploratory analysis indicated that groups with higher average cortisol levels also had greater variation in this response variable. To make the variation more uniform, we used a logarithmic transform of each fish's cortisol per unit weight as the dependent variable in our analyses. This action made the assumptions of normality and homoscedasticity (standard deviations were equal) of our analyses reasonable. [96]

Item 8. Experimental animals

8a. Provide details of the animals used, including species, strain and substrain, sex, age or developmental stage, and weight.

Explanation

The species, strain, substrain, sex, weight, and age of animals are critical factors that can influence most experimental results [97-101]. Reporting the characteristics of all animals used is equivalent to standardised human patient demographic data; these data support both the internal and external validity of the study results. It enables other researchers to repeat the experiment and generalise the findings. It also enables readers to assess whether the animal characteristics chosen for the experiment are relevant to the research objectives.

Report age and weight for each group, include summary statistics (e.g. mean and standard deviation) and, if possible, baseline values for individual animals (e.g. as supplementary information or a link to a publicly accessible data repository). For most species, precise reporting of age is more informative than a description of the developmental status (e.g. a mouse referred to as an adult can vary in age from six to 20 weeks [102]). In some cases, however, reporting the developmental stage is more informative than chronological age, for example in juvenile *Xenopus*, where rate of development can be manipulated by incubation temperature [103].

Example

1) "One hundred and nineteen male mice were used: C57BL/6OlaHsd mice (n = 59), and BALB/c OlaHsd mice (n = 60) (both from Harlan, Horst, The Netherlands). At the time of the EPM test the mice were 13 weeks old and had body weights of 27.4 ± 0.4 g and 27.8 ± 0.3 g, respectively (mean ± SEM)." [104]
2) "Histone Methylation Profiles and the Transcriptome of X. tropicalis Gastrula Embryos. To generate epigenetic profiles, ChIP was performed using specific antibodies against trimethylated H3K4 and H3K27 in Xenopus gastrula-stage embryos (Nieuwkoop-Faber stage 11–12), followed by deep sequencing (ChIP-seq). In addition, polyA-selected RNA (stages 10–13) was reverse transcribed and sequenced (RNA-seq)." [105]

8b. Provide further relevant information on the provenance of animals, health/immune status, genetic modification status, genotype, and any previous procedures.

Explanation

The animals' provenance, their health or immune status and their history of previous testing or procedures, can influence their physiology and behaviour as well as their response to treatments, and thus impact on study outcomes. For example, animals of the same strain, but from different sources, or animals obtained from the same source but at different times, may be genetically different [17]. The immune or microbiological status of the animals can also influence welfare, experimental variability and scientific outcomes [106-108].

Report the health status of animals in the study, and any previous procedures the animals have undergone. For genetically modified animals, describe the genetic modification status (e.g. knockout, overexpression), genotype (e.g. homozygous, heterozygous), manipulated gene(s), genetic methods and technologies used to generate the animals, how the genetic modification was confirmed, and details of animals used as controls (e.g. littermate controls [109]).

Reporting the correct nomenclature is crucial to understanding the data and ensuring that the research is discoverable and replicable [110-112]. Useful resources for reporting nomenclature for different species include:

- Mice - International Committee on Standardized Genetic Nomenclature (https://www.jax.org/jax-mice-and-services/customer-support/technical-support/genetics-and-nomenclature)
- Rats - Rat Genome and Nomenclature Committee (https://rgd.mcw.edu/)
- Zebrafish - Zebrafish Information Network (http://zfin.org/)
- Xenopus - Xenbase (http://www.xenbase.org/entry/)
- Drosophila – FlyBase (http://flybase.org/)

Examples

1) "A construct was engineered for knockin of the miR-128 (miR-128-3p) gene into the Rosa26 locus. Rosa26 genomic DNA fragments (~1.1 kb and ~4.3 kb 5' and 3' homology arms, respectively) were amplified from C57BL/6 BAC DNA, cloned into the pBasicLNeoL vector sequentially by in-fusion cloning, and confirmed by sequencing. The miR-128 gene, under the control of tetO-minimum promoter, was also cloned into the vector between the two homology arms. In addition, the targeting construct also contained a loxP sites flanking the neomycin resistance gene cassette for positive selection and a diphtheria toxin A (DTA) cassette for negative selection. The construct was linearized with ClaI and electroporated into C57BL/6N ES cells. After G418 selection, seven-positive clones were identified from 121 G418-resistant clones by PCR screening. Six-positive clones were expanded and further analyzed by Southern blot analysis, among which four clones were confirmed with correct targeting with single-copy integration. Correctly targeted ES cell clones were injected into blastocysts, and the blastocysts were implanted into pseudo-pregnant mice to generate chimeras by Cyagen Biosciences Inc. Chimeric males were bred with Cre deleted mice from Jackson Laboratories to generate neomycin-free knockin mice. The

correct insertion of the miR-128 cassette and successful removal of the neomycin cassette were confirmed by PCR analysis with the primers listed in Supplementary Table 1." [113]
2) "The C57BL/6J (Jackson) mice were supplied by Charles River Laboratories. The C57BL/6JOlaHsd (Harlan) mice were supplied by Harlan. The α-synuclein knockout mice were kindly supplied by Prof. [X] (Cardiff University, Cardiff, United Kingdom.) and were congenic C57BL/6JCrl (backcrossed for 12 generations). TNFα−/− mice were kindly supplied by Dr. [Y] (Queens University, Belfast, Northern Ireland) and were inbred on a homozygous C57BL/6J strain originally sourced from Bantin & Kingman and generated by targeting C57BL/6 ES cells. T286A mice were obtained from Prof. [Z] (University of California, Los Angeles, CA). These mice were originally congenic C57BL/6J (backcrossed for five generations) and were then inbred (cousin matings) over 14 y, during which time they were outbred with C57BL/6JOlaHsd mice on three separate occasions." [114]

Item 9. Experimental procedures

For each experimental group, including controls, describe the procedures in enough detail to allow others to replicate them, including:

9a. What was done, how it was done and what was used.

Explanation

Essential information to describe in the manuscript includes the procedures used to develop the model (e.g. induction of the pathology), the procedures used to measure the outcomes, and pre- and post-experimental procedures, including animal handling and welfare monitoring. Animal handling can be a source of stress and the specific method used (e.g. mice picked up by tail or in cupped hands) can affect research outcomes [56, 115, 116]. Details about animal care and monitoring intrinsic to the procedure are discussed in further detail in **item 16 – Animal care and monitoring**. Provide enough detail to enable others to replicate the methods and highlight any quality assurance and quality control used [117, 118]. A schematic of the experimental procedures with a timeline can give a clear overview of how the study was conducted. Information relevant to distinct types of interventions and resources are described in Box 6.

Box 6: Examples of information to include when reporting specific types experimental procedures and resources

Procedures	**Resources**
<u>Pharmacological procedures (intervention and control)</u> ▪ Drug formulation ▪ Dose ▪ Volume ▪ Concentration ▪ Site and route of administration ▪ Frequency of administration ▪ Vehicle or carrier solution formulation and volume ▪ Any evidence that the pharmacological agent used reaches the target tissue <u>Surgical procedures (including sham surgery)</u> ▪ Description of the surgical procedure ▪ Anaesthetic used (including dose and other information listed in pharmacological procedures section above) ▪ Pre and post analgesia regimen ▪ Pre-surgery procedures (e.g. fasting) ▪ Aseptic techniques ▪ Monitoring (e.g. assessment of surgical anaesthetic plane)	<u>Cell lines</u> ▪ Identification ▪ Provenance ▪ Verification and authentication ▪ RRID [120, 121] <u>Reagents (e.g. antibodies, chemicals)</u> ▪ Manufacturer ▪ Supplier ▪ Catalogue number ▪ Lot number (if applicable) ▪ RRID <u>Equipment and software</u> ▪ Manufacturer ▪ Supplier ▪ Model/version number ▪ Calibration procedures (if applicable) ▪ RRID

▪ Whether the procedure is terminal or not ▪ Post-surgery procedures ▪ Duration of the procedure and duration of anaesthesia ▪ Physical variables measured <u>Pathogen infection (intervention and control)</u> ▪ Infectious agent ▪ Dose load ▪ Vehicle or carrier solution formulation and volume ▪ Site and route of infection ▪ Timing or frequency of infection <u>Euthanasia</u> ▪ Method of euthanasia, including the humane standards the method complies with, such as AVMA [119] ▪ Pharmacological agent, if used (including dose and information listed in pharmacological procedures section above) ▪ Any measures taken to reduce pain and distress before or during euthanasia ▪ Timing of euthanasia ▪ Tissues collected post-euthanasia and timing of collection	

Where available, cite the Research Resource Identifier (RRID) for reagents and tools used [120, 121]. RRIDs are unique and stable, allowing unambiguous identification of reagents or tools used in a study, aiding other researchers to replicate the methods.

Detailed step-by-step procedures can also be saved and shared online, for example using Protocols.io [122], which assigns a DOI to the protocol and allows cross-referencing between protocols and publications.

Examples

1) "Fig… shows the timeline for instrumentation, stabilization, shock/injury, and resuscitation…. Animals were food-deprived for 18 hours before surgery, but allowed free access to water. On the morning of surgery, swine were sedated with tiletamine-zolazepam (Telazol®; 5-8 mg/kg IM; Zoetis Inc., Kalamazoo MI) in the holding pen, weighed, and masked with isoflurane (3%, balance 100% O2) for transport to the lab. The marginal ear vein was catheterized for administration of atropine (0.02 mg/kg IV; Sparhawk Laboratories, Lenexa KS), and buprenorphine for pre-emptive analgesia (3 mg/ml IV; ZooPharm, Laramie WY). Ophthalmic ointment (Puralube®; Fera Pharmaceuticals) was applied to prevent corneal drying. Animals were intubated in dorsal recumbency with a cuffed 6 or 7 Fr endotracheal tube. Anesthetic plane was maintained by isoflurane (1-1.5%; 21-23 % O2, balance N2). Oxygen saturation (sPO2, %) and heart rate (HR) were monitored with a veterinary pulse oximeter placed in the buccal cavity (Masimo SET Radical-7; Irvine CA). Core temperature was monitored with a rectal probe and maintained at 36.5-38°C with a microprocessor-controlled feedback water blanket (Blanketrol® II, Cincinnati Sub-Zero (CSZ) Cincinnati, OH) placed under the animal. Anesthetic depth was assessed every 5 min for the duration of the experiment by reflexes (corneal touch, pedal flexion, coronary band pinch) and vital signs (sPO2, HR, core temperature). *Fig.... Experimental time line for instrumentation, shock/injury, resuscitation, post-resuscitation monitoring, and blood sampling in a swine model of polytrauma and coagulopathy." [123]*
2) "For the diet-induced obesity (DIO) model, eight-week-old male mice had ad libitum access to drinking water and were kept on standard chow (SFD, 10.9 kJ/g) or on western high-fat diet (HFD; 22 kJ/g; kcal from 42% fat, 43% from carbohydrates and 15% from protein; E15721-34, Ssniff, Soest, Germany) for 15 weeks (http://dx.doi.org/10.17504/protocols.io.kbacsie)." [124]

9b. When and how often.

Explanation

Clearly report the frequency and timing of experimental procedures and measurements, including the light and dark cycle (e.g. 12:12), circadian time cues (e.g. lights on at 08:00), and experimental time sequence (e.g. interval between baseline and comparator measurements or interval between procedures and measurements). Along with innate circadian rhythms, these can affect research outcomes such as behavioural, physiological, and immunological parameters [125, 126]. Also report the timing and frequency of welfare assessments, taking into consideration the normal activity patterns (see **item 16 – Animal care and monitoring**). For example, nocturnal animals may not show behavioural signs of discomfort during the day [127].

Examples

1) "Blood pressure, heart rate, oxygen saturation and amount of blood extracted were recorded every 5 minutes. Blood samples were drawn at baseline (pre injury), 0 minutes (immediately after injury), and after 30 and 60 minutes." [128]
2) "After a 5-h fast (7:30–12:30am), awake and freely moving mice were randomized and subjected to three consecutive clamps performed in the same mice as described above, with a 2 days recovery after each hyperinsulinemic/hypoglycemic (mHypo, n = 6) or hyperinsulinemic/euglycemic (mEugly, n = 4) clamps." [129]

9c. Where (including detail of any acclimation periods).

Explanation

Physiological acclimation after a stressful event, such as transport (e.g. between supplier, animal facility and laboratory), but before the experiment begins allows stabilisation of physiological responses of the animal [130, 131]. Protocols vary depending on species, strain, and outcome; for example physiological acclimation following transportation of different animals can take anywhere from 24 hours to more than one week [132]. Procedural acclimation, immediately before a procedure, allows stabilisation of the animals' responses after unaccustomed handling, novel environments, and previous procedures, which otherwise can induce behavioural and physiological changes [133, 134].

Indicate where studies were performed (e.g. dedicated laboratory space or animal facility, home cage, open field arena, water maze) and whether periods of physiological or procedural acclimation were included in the study protocol, including type and duration. If the study involved multiple sites explicitly state where each experiment and sample analysis was performed. Include any accreditation of laboratories if appropriate (e.g. if samples are sent to a commercial laboratory for analysis).

Example

1) "Fish were singly housed for 1 week before being habituated to the conditioning tank over 2 consecutive days. The conditioning tank consisted of an opaque tank measuring 20 cm (w) 15 cm (h) 30 cm (l) containing 2.5 l of aquarium water with distinct visual cues (spots or stripes) on walls at each end of the tank… During habituation, each individual fish was placed in the conditioning apparatus for 20 minutes with free access to both compartments and then returned to its home tank." [135]

9d. Why (provide rationale for procedures).

Explanation

There may be numerous approaches to investigate any given research problem, therefore it is important to explain why a particular procedure or technique was chosen. This is especially relevant when procedures are novel or

specific to a research laboratory, or constrained by the animal model or experimental equipment (e.g. route of administration determined by animal size [136]).

Examples

1) "Because of the very small caliber of the murine tail veins, partial paravenous injection is common if 18F-FDG is administered by tail vein injection (intravenous). This could have significantly biased our comparison of the biodistribution of 18F-FDG under various conditions. Therefore, we used intraperitoneal injection of 18F-FDG for our experiments evaluating the influence of animal handling on 18F-FDG biodistribution." [137]

2)"Since Xenopus oocytes have a higher potential for homologous recombination than fertilized embryos (Hagmann et al., 1996), we next tested whether the host transfer method could be used for efficient HDR-mediated knock-in. We targeted the C-terminus of X. laevis Ctnnb1 (β-catenin), a key cytoskeletal protein and effector of the canonical Wnt pathway, because previous studies have shown that addition of epitope tags to the C-terminus do not affect the function of the resulting fusion protein (Fig. 2A) (Evans et al., 2010; Miller and Moon, 1997). CRISPR components were injected into X. laevis oocytes followed by host transfer or into embryos." [138]

Item 10. Results

For each experiment conducted, including independent replications, report:

10a. Summary/descriptive statistics for each experimental group, with a measure of variability where applicable.

Explanation

Summary/descriptive statistics provide a quick and simple description of the data, they communicate quantitative results easily and facilitate visual presentation. For continuous data, these descriptors include a measure of central tendency (e.g. mean, median) and a measure of variability (e.g. quartiles, range, standard deviation) to help readers assess the precision of the data collected. Categorical data can be expressed as counts, frequencies, or proportions.

Report data for all experiments conducted. If a complete experiment is repeated on a different day, or under different conditions, report the results of all repeats, rather than selecting data from representative experiments. Report the exact number of experimental units per group so readers can gauge reliability of results (see **item 2 – Sample size**, and **item 3 – Inclusion and exclusion criteria**). Present data clearly as text, in tables, or in graphs, to enable information to be evaluated, or extracted for future meta-analyses. Report descriptive statistics with a clearly identified measure of variability for each group. Example 1 shows data summarised as medians and, in brackets, lower and upper quartiles. Boxplots are a convenient way to summarise continuous data, plotted as median, interquartile range, minimum, maximum and outliers, as shown in Example 2.

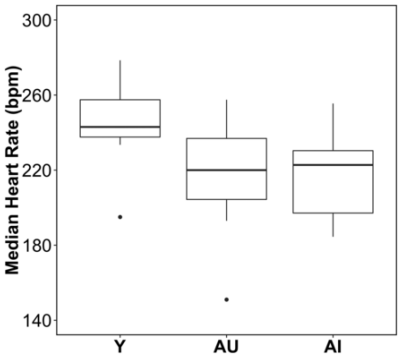
Examples

1) "Features of autism compared between the control group and the autistic model groups" [139]

	Median (quartiles)	
	Control group (n=13)	*Autism model group (n=14)*
Birth weight (g)	*4.2 (3.9 – 4.5)*	*3.6 (2.9 – 5.0)*
Weight at 28 weeks (g)	*47.8 (44.0 – 51.8)*	*32.5 (29.0 – 53.5)*
Number of squares traversed	*6.0 (4.0 – 11.0)*	*3.0 (2.0 – 7.0)*
Rearing movements	*6.0 (4.0 – 13.0)*	*3.0 (2.0 – 6.0)*
Grooming movements (number of rats performing grooming movements/total in the group)	*4/13*	*2/14*

[139]

2) "Boxplots of median heart rate (beats per minute, bpm) during rs-fMRI scans for Y (n = 12), AU (n = 12), and AI (n = 12) animals."[140]



10b. If applicable, the effect size with a confidence interval.

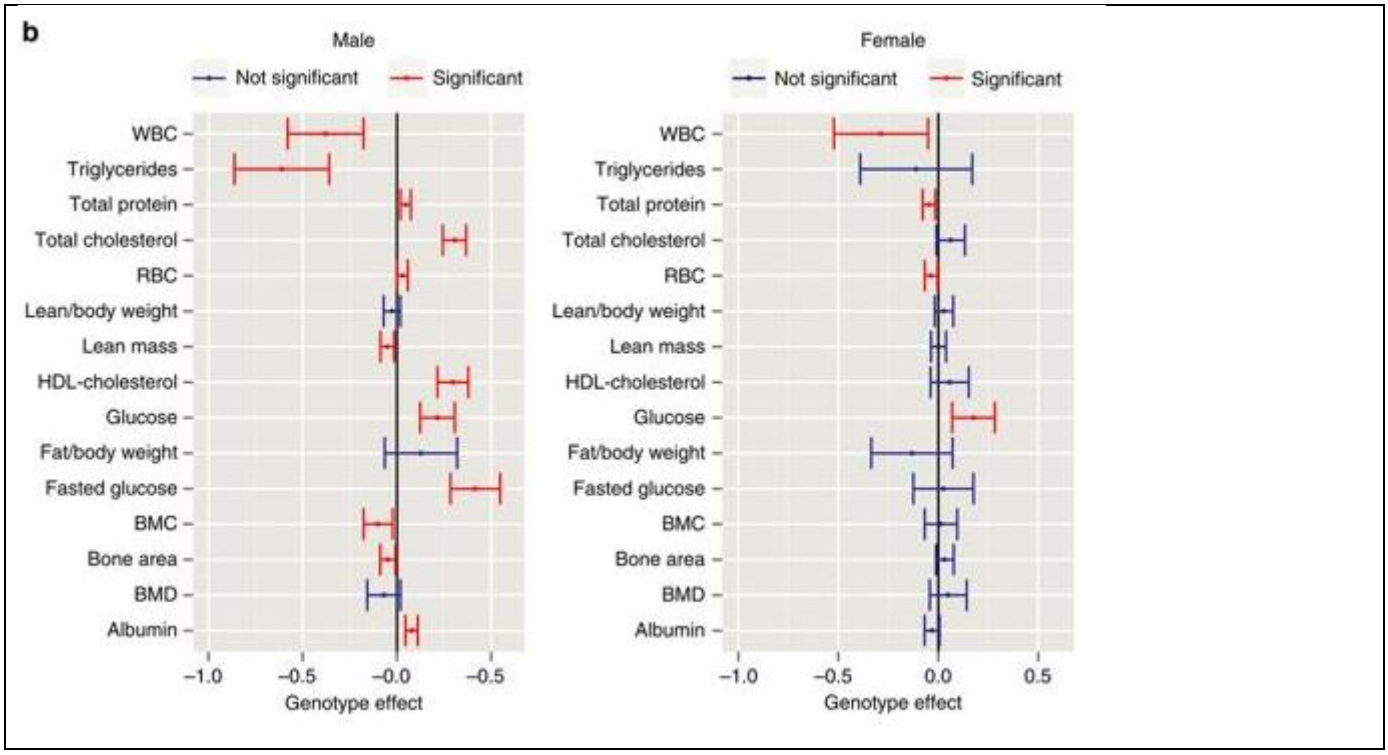
Explanation

An effect size is a quantitative measure that estimates the magnitude of differences between groups, or relationships between variables. It can be used to assess the patterns in the data collected and make inferences about the wider population from which the sample came. The confidence interval for the effect indicates how precisely the effect has been estimated, and tells the reader about the strength of the effect [141]. For example, if zero is included in the 95% confidence interval, the presence of an effect cannot be assumed. In studies where statistical power is low, and/or hypothesis-testing is inappropriate, providing the effect size and confidence interval indicates how small or large an effect might really be, so a reader can judge the biological significance of the data [142, 143]. Reporting effect sizes with confidence intervals also facilitates extraction of useful data for systematic review and meta-analysis. Where multiple independent studies included in a meta-analysis show quantitatively similar effects, even if each is statistically non-significant, this provides powerful evidence that a relationship is 'real', although small.

Report all analyses performed, even those providing non-statistically significant results. Report the effect size, to indicate the size of the difference between groups in the study, with a confidence interval, to indicate the precision of the effect size estimate.

Example

1) "For all traits identified as having a significant genotype effect for the Usp47tm1b(EUCOMM)Wtsi line (MGI:5605792), a comparison is presented of the standardized genotype effect with 95% confidence interval for each sex with no multiple comparisons correction. Standardization, to allow comparison across variables, was achieved by dividing the genotype estimate by the signal seen in the wildtype population. Shown in red are statistically significant estimates. RBC: red blood cells; BMC: bone mineral content; BMD: bone mineral density; WBC: white blood cells." [28]



2. Recommended Set

The Recommended Set (Box 7) adds context to the study described, including further detail about the methodology and advice on what to include in the more narrative parts of a manuscript. Items are presented in a logical order, there is no ranking within the set.

Box 7: ARRIVE Recommended Set

11. Abstract
12. Background
13. Objectives
14. Ethical statement
15. Housing and husbandry
16. Animal care and monitoring
17. Interpretation /scientific implications
18. Generalisability /translation
19. Protocol registration
20. Data access
21. Declaration of interests

Item 11. Abstract

Provide an accurate summary of the research objectives, animal species, strain and sex, key methods, principal findings, and study conclusions.

Explanation

A transparent and accurate abstract increases the utility and impact of the manuscript, and allows readers to assess the reliability of the study [144]. The abstract is often used as a screening tool by readers to decide whether to read the full article or to select an article for inclusion in a systematic review. However, abstracts often either do not contain enough information for this purpose [145, 146], or contain information that is inconsistent with the results in the rest of the manuscript [147, 148].

To maximise utility, include details of the species, sex and strain of animals used, and accurately report the methods, results and conclusions of the study. Also describe the objectives of the study, including whether it was designed to either test a specific hypothesis or to generate a new hypothesis (see **Item 13 – Objectives**). Incorporating this information will enable readers to interpret the strength of evidence, and judge how the study fits within the wider knowledge base.

Examples

1) "BACKGROUND AND PURPOSE: Asthma is an inflammatory disease that involves airway hyperresponsiveness and remodelling. Flavonoids have been associated to anti-inflammatory and antioxidant activities and may represent a potential therapeutic treatment of asthma. Our aim was to evaluate the effects of the sakuranetin treatment in several aspects of experimental asthma model in mice.

EXPERIMENTAL APPROACH: Male BALB/c mice received ovalbumin (i.p.) on days 0 and 14, and were challenged with aerolized ovalbumin 1% on days 24, 26 and 28. Ovalbumin-sensitized animals received vehicle (saline and dimethyl sulfoxide, DMSO), sakuranetin (20 mg kg–1per mice) or dexamethasone (5 mg kg–1 per mice) daily beginning from 24th to29th day. Control group received saline inhalation and nasal drop vehicle. On day 29, we determined the airway hyperresponsiveness, inflammation and remodelling as well as specific IgE antibody. RANTES, IL-5, IL-4, Eotaxin, IL-10, TNF-a, IFN-g and GMC-SF content in lung homogenate was performed by Bioplex assay, and 8-isoprostane and NF-kB activations were visualized in inflammatory cells by immunohistochemistry.

KEY RESULTS: We have demonstrated that sakuranetin treatment attenuated airway hyperresponsiveness, inflammation and remodelling; and these effects could be attributed to Th2 pro-inflammatory cytokines and oxidative stress reduction as well as control of NF-kB activation.

CONCLUSIONS AND IMPLICATIONS: These results highlighted the importance of counteracting oxidative stress by flavonoids in this asthma model and suggest sakuranetin as a potential candidate for studies of treatment of asthma." [149]

2) "In some parts of the world, the laboratory pig (Sus scrofa) is often housed in individual, sterile housing which may impose stress. Our objectives were to determine the effects of isolation and enrichment on pigs housed within the PigTurn® — a novel penning system with automated blood sampling — and to investigate tear staining as a novel welfare indicator. Twenty Yorkshire × Landrace weaner pigs were randomly assigned to one of four treatments in a 2 × 2 factorial combination of enrichment (non-enriched [NE] or enriched [E]) and isolation (visually isolated [I] or able to see another pig [NI]). Pigs were catheterised and placed into the PigTurns® 48 h post recovery. Blood was collected automatically twice daily to determine white blood cell (WBC) differential counts and assayed for cortisol. Photographs of the eyes were taken daily and tear staining was quantified using a 0–5 scoring scale and Image-J software to measure stain area and perimeter. Behaviour was video recorded and scan sampled to determine time budgets. Data were analysed as an REML using the MIXED procedure of SAS. Enrichment tended to increase proportion of time standing and lying laterally and decrease plasma cortisol, tear-stain area and perimeter. There was a significant isolation by enrichment interaction. Enrichment given to pigs housed in isolation had no effect on plasma cortisol, but greatly reduced it in non-isolated pigs. Tear-staining area and perimeter were highest in the NE-I treatment compared to the other three treatments. Eosinophil count was highest in the E-NI treatment and lowest in the NE-I treatment. The results suggest that in the absence of enrichment, being able to see another animal but not interact may be frustrating. The combination of no enrichment and isolation maximally impacted tear staining and eosinophil numbers. However, appropriate enrichment coupled with proximity of another pig would appear to improve welfare." [150]

Item 12. Background

12a. Include sufficient scientific background to understand the rationale and context for the study, and explain the experimental approach.

Explanation

Scientific background information for an animal study should demonstrate a clear evidence gap and explain why an *in vivo* approach was warranted. Systematic reviews of the animal literature provide the most convincing evidence that a research question has not been conclusively addressed, by showing the extent of current evidence within a field of research. They can also inform the choice of animal model by providing a comprehensive overview of the models used along with their benefits and limitations [151, 152].

Describe the rationale and context of the study and how it relates to other research, including relevant references to previous work. Outline evidence underpinning the hypothesis or objectives and explain why the experimental approach is best suited to answer the research question.

Example

1) "For decades, cardiovascular disease has remained the leading cause of mortality worldwide ...[and] cardiovascular research has been performed using healthy and young, non-diseased animal models. Recent failures of cardioprotective therapies in obese insulin-resistant, diabetic, metabolic syndrome-affected and aged animals that were otherwise successful in healthy animal models has highlighted the need for the development of animal models of disease that are representative of human clinical conditions… In the clinical setting, elderly male patients often present with both testosterone deficiency (TD) and the metabolic syndrome (MetS). A strong and compounding association exists between MetS and TD which may have significant impact on cardiovascular disease and its outcomes which is not addressed by current models…. their mutual presentation in the clinical setting warrants the development of appropriate animal models of the MetS with hypogonadism, especially in the context of cardiovascular disease research." [153]

12b. Explain how the animal species and model used address the scientific objectives and, where appropriate, the relevance to human biology.

Explanation

Provide enough detail for the reader to assess the suitability of the animal model used to address the research question. Include information on the rationale for choosing a particular species, explain how the outcome measures assessed are relevant to the condition under study, and how the model was validated. Stating that an animal model is commonly used in the field is not appropriate, and a well-considered, detailed rationale should be provided.

When the study models a human disease, indicate how the model is appropriate for addressing the specific objectives of the study [154]. This can include a description of how the induction of the disease, disorder, or injury is sufficiently analogous to the human condition, how the model responds to known clinically-effective treatments, how similar symptoms are to the clinical disease and how animal characteristics were selected to represent the age, sex, and health status of the clinical population [14].

Examples

1) "… we selected a pilocarpine model of epilepsy that is characterized by robust, frequent spontaneous seizures acquired after a brain insult, well-described behavioral abnormalities, and poor responses to antiepileptic drugs. These animals recapitulate several key features of human temporal lobe epilepsy, the most common type of epilepsy in adults." [155]
2) "Transplantation of healthy haematopoietic stem cells (HSCs) is a critical therapy for a wide range of malignant haematological and non-malignant disorders and immune dysfunction (Snowden et al., 2012; Sykes & Nikolic, 2005; Thomas et al., 1957)…. Zebrafish are already established as a successful model to study the haematopoietic system, with significant homology with mammals (de Jong & Zon, 2005; Gering & Patient, 2005; Kissa & Herbomel, 2010; Renshaw & Trede, 2012; Traver et al., 2003; White et al., 2008). Imaging of zebrafish transparent embryos remains a powerful tool and has been critical to confirm that the zebrafish Caudal

Haematopoietic Tissue (CHT) is comparable to the mammalian foetal haematopoietic niche (Gering & Patient, 2005; Kissa & Herbomel, 2010; Tamplin et al., 2015). Xenotransplantation in zebrafish embryos has revealed highly conserved mechanisms between zebrafish and mammals. Recently, murine bone marrow cells were successfully transplanted into zebrafish embryos, revealing highly conserved mechanism of haematopoiesis between zebrafish and mammals (Parada-Kusz et al., 2017). Additionally, CD34 enriched human cells transplanted into zebrafish were shown to home to the CHT and respond to zebrafish stromal-cell derived factors (Staal et al., 2016).” [156]

Item 13. Objectives

Clearly describe the research question, research objectives and, where appropriate, specific hypotheses being tested.

Explanation

Explaining the purpose of the study by describing the question(s) that the research addresses, allows readers to determine if the study is relevant to them. Readers can also assess the relevance of the model organism, procedures, outcomes measured, and analysis used.

Knowing if a study is exploratory or hypothesis-testing is critical to its interpretation. A typical exploratory study may measure multiple outcomes and look for patterns in the data, or relationships that can be used to generate hypotheses. It may also be a pilot study which aims to inform the design or feasibility of larger subsequent experiments. Exploratory research helps researchers to design hypothesis-testing experiments, by choosing what variables or outcome measures to focus on in subsequent studies.

Testing a specific hypothesis has implications for both the study design and the data analysis [17, 157]. For example, an experiment designed to detect a hypothesised effect will likely need to be analysed with inferential statistics, and a statistical estimation of the sample size will need to be performed *a priori* (see **item 2 – Sample size**). Hypothesis-testing studies also have a pre-defined primary outcome measure, which is used to assess the evidence in support of the specific research question (see **item 6 – Outcome measures**).

In contrast, exploratory research investigates many possible effects, and is likely to yield more false positive results because some will be positive by chance; thus results from well-designed hypothesis-testing studies provide stronger evidence. Independent replication and meta-analysis can further increase the confidence in conclusions.

Clearly outline the objective(s) of the study, including whether it is hypothesis-testing or exploratory, or if it includes research of both types. Hypothesis-testing studies may collect additional information for exploratory purposes, it is important to distinguish which hypotheses were prespecified and which originated after data inspection, especially when reporting unanticipated effects or outcomes that were not part of the original study design.

Examples

1) “The primary objective of this study was to investigate the cellular immune response to MSC injected into the striatum of allogeneic recipients (6-hydroxydopamine [6-OHDA]-hemilesioned rats, an animal model of Parkinson's disease [PD]), and the secondary objective was to determine the ability of these cells to prevent nigrostriatal dopamine depletion and associated motor deficits in these animals.” [158]
2) “In this exploratory study, we aimed to investigate whether calcium electroporation could initiate an anticancer immune response similar to electrochemotherapy. To this end, we treated immunocompetent balb/c mice with CT26 colon tumors with calcium electroporation, electrochemotherapy, or ultrasound-based delivery of calcium or bleomycin.” [159]

Item 14. Ethical statement

Provide the name of the ethical review committee or equivalent that has approved the use of animals in this study and any relevant licence or protocol numbers (if applicable). If ethical approval was not sought or granted, provide a justification.

Explanation

Authors are responsible for complying with regulations and guidelines relating to the use of animals for scientific purposes. This includes ensuring that they have the relevant approval for their study from an appropriate ethics committee and/or regulatory body before the work starts. The ethical statement provides editors, reviewers and readers with assurance that studies have received this ethical oversight [12]. This also promotes transparency and understanding about the use of animals in research and fosters public trust.

Provide a clear statement explaining how the study conforms to appropriate regulations and guidelines, and indicate relevant licence/protocol numbers so that the study can be identified. Add also any relevant accreditation e.g. AAALAC (American Association for Accreditation of Laboratory Animal Care) [160] or GLP (Good Laboratory Practice).

If the research is not covered by any regulation and formal ethical approval is not required (e.g. a study using animal species not protected by regulations or law), demonstrate that international standards were complied with and cite the appropriate reference. In such cases, provide a clear statement explaining why the research is exempt from regulatory approval.

Examples

1) "All procedures were conducted in accordance with the United Kingdom Animal (Scientific Procedures) Act 1986, approved by institutional ethical review committees (Alderley Park Animal Welfare and Ethical Review Board and Babraham Institute Animal Welfare and Ethical Review Board) and conducted under the authority of the Project Licence (40/3729 and 70/8307, respectively)." [161]
2) "All protocols in this study were approved by the Committee on the Ethics of Animal Experiments of Fuwai Hospital, Peking Union Medical College and the Beijing Council on Animal Care, Beijing, China (IACUC permit number: FW2010-101523), in compliance with the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH publication no.85-23, revised 1996)." [162]
3) "Samples and data were collected according to Institut de Sélection d'Animale (ISA) protocols, under the supervision of ISA employees. Samples and data were collected as part of routine animal data collection in a commercial breeding program for layer chickens in The Netherlands. Samples and data were collected on a breeding nucleus of ISA for breeding purposes only, and is a non-experimental, agricultural practice, regulated by the Act Animals, and the Royal Decree on Procedures. The Dutch Experiments on Animals Act does not apply to non-experimental, agricultural practices. An ethical review by the Statement Animal Experiment Committee was therefore not required. No extra animal discomfort was caused for sample collection for the purpose of this study." [163]

Item 15. Housing and husbandry

Provide details of housing and husbandry conditions, including any environmental enrichment.

Explanation

The environment determines the health and wellbeing of the animals and every aspect of it can potentially affect their behavioural and physiological responses, thereby affecting research outcomes. Different studies may be sensitive to different environmental factors, and particular aspects of the environment necessary to report may depend on the type of study. Examples of housing and husbandry conditions known to affect animal welfare and research outcomes are listed in Table 1; consider reporting these elements and any other housing and husbandry conditions likely to influence the study outcomes.

Table 1: Examples of information to consider when reporting housing and husbandry, and their effects on laboratory animals

Information to report	**Examples of effects on laboratory animals**
Cage/housing system (type and dimensions)	Affects behaviour [164] and fear learning [165]. Tank colour affects stress in *Xenopus laevis* [166].
Food and water (type, composition and access)	Affects body weight, tumour development, nephropathy severity [167], and the threshold for developing Parkinsonian symptoms [168]. Maternal diet affects offspring body weight [169].
Bedding and nesting material	Affects behavioural responses to stress [170] and pain [171].
Temperature and humidity	Modifies tumour progression [172].
Sanitation (frequency of cage changes, material transferred)	Affects blood pressure, heart rate, behaviour [173]. Adds an additional source of variation [174, 175].
Social environment (group size and composition)	Compromises animal welfare [176]. Induces aggressive behaviour [177].
Biosecurity (level)	The microbiological status of animals causes variation in systemic disease parameters [178].
Lighting (type, schedule and intensity)	Modifies immune and stress responses[179].
Environmental enrichment	Reduces anxiety [180, 181], stress [180, 181] and abnormal repetitive behaviour [182-184]. Reduces susceptibility to epilepsy [185] and osteoarthritis [186] and modifies the pathology of neurological disorders [187].
Sex of the experimenter	Affects physiological stress and pain behaviour [188].

Environment, either deprived or enriched, can affect a wide range of physiological and behavioural responses [189]. Specific details to report include, but are not limited to, structural enrichment (e.g. elevated surfaces, dividers), resources for species-typical activities (e.g. nesting material, shelters, gnawing sticks), toys and or other tools used to stimulate exploration, exercise (e.g. running wheel), and novelty. If no environmental enrichment was provided, this should be clearly stated with justification. Similarly, scientific justification needs to be reported for withholding food and water [190], and for singly housing animals [191, 192].

Examples

1) "Breeding colonies were kept in individually ventilated cages (IVCs; Tecniplast, Italy) at a temperature of 20°C to 24°C, humidity of 50% to 60%, 60 air exchanges per hour in the cages, and a 12/12-hour light/dark cycle with the lights on at 5:30 AM. The maximum caging density was five mice from the same litter and sex starting from weaning. As bedding, spruce wood shavings (Lignocel FS-14; J. Rettenmaier und Soehne GmbH, Rosenberg, Germany) were provided. Mice were fed a standardized mouse diet (1314, Altromin, Germany) and provided drinking water ad libitum. All materials, including IVCs, lids, feeders, bottles, bedding, and water were autoclaved before use. Sentinel mice were negative for at least all Federation of laboratory animal science associations (FELASA)-relevant murine infectious agents [30] as diagnosed by our health monitoring laboratory, mfd Diagnostics GmbH, Wendelsheim, Germany." [193]
2) "Same sex litter mates were housed together in individually ventilated cages with two or four mice per cage. All mice were maintained on a regular diurnal lighting cycle (12:12 light:dark) with ad libitum access to food (7012 Harlan Teklad LM-485 Mouse/Rat Sterilizable Diet) and water. Chopped corn cob was used as bedding. Environmental enrichment included nesting material (Nestlets, Ancare, Bellmore, NY, USA), PVC pipe, and shelter (Refuge XKA-2450-087, Ketchum Manufacturing Inc., Brockville, Ontario, Canada). Mice were housed under

broken barrier-specific pathogen-free conditions in the Transgenic Mouse Core Facility of Cornell University, accredited by AAALAC (The Association for Assessment and Accreditation of Laboratory Animal Care International)." [194]

Item 16. Animal care and monitoring

16a. Describe any interventions or steps taken in the experimental protocols to reduce pain, suffering and distress.

Explanation

A safe and effective analgesic plan is critical to relieve pain, suffering and distress. Untreated pain can affect the animals' biology and add variability to the experiment; however specific pain management procedures can also introduce variability, affecting experimental data [195, 196]. Under-reporting of welfare management procedures can also contribute to the perpetuation of non-compliant methodologies and insufficient or inappropriate use of analgesia [196] or other welfare measures. A thorough description of the procedures used to alleviate pain, suffering and distress provides practical information for researchers to replicate the method.

Clearly describe pain management strategies, including:

- specific analgesic
- administration method (formulation, route, dose, concentration, volume, frequency, timing, and equipment used)
- rationale for the choice (e.g. animal model, disease/pathology, procedure, mechanism of action, pharmacokinetics, personnel safety)
- protocol modifications to reduce pain, suffering and distress (e.g. changes to the anaesthetic protocol, increased frequency of monitoring, procedural modifications, habituation, etc.)

If analgesics or other welfare measures, reasonably expected for the procedure performed, are not performed for experimental reasons, report the scientific justification [197].

Examples

1) "If piglets developed diarrhea, they were placed on an electrolyte solution and provided supplemental water, and if the diarrhea did not resolve within 48 h, piglets received a single dose of ceftiofur (5.0 mg ceftiofur equivalent/kg of body weight i.m [Excede, Zoetis, Florham Park, NJ]). If fluid loss continued after treatment, piglets then received a single dose of sulfamethoxazole and trimethoprim oral suspension (50 mg/8 mg per mL, Hi-Tech Pharmacal, Amityville, NY) for 3 consecutive days." [198]
2) "One hour before surgery, we administered analgesia to the mice by offering them nut paste (Nutella; Ferrero, Pino Torinese, Italy) containing 1 mg per kg body weight buprenorphine (Temgesic; Schering-Plough Europe, Brussels, Belgium) for voluntary ingestion, as described previously. The mice had been habituated to pure nut paste for 2 d prior to surgery." [199]
3) "If a GCPS score equal or greater than 6 (out of 24) was assigned postoperatively, additional analgesia was provided with methadone 0.1 mg kg−1 IM (or IV if required), and pain reassessed 30 minutes later. The number of methadone doses was recorded." [44]

16b. Report any expected or unexpected adverse events.

Explanation

Reporting adverse events allows other researchers to minimise the risk of these events occurring in their own studies and to plan appropriate welfare assessments. If the experiment is testing the efficacy of a treatment, the occurrence of adverse events may alter the balance between treatment benefit and risk [33].

Report any adverse events, expected or unexpected, that had a negative impact on the welfare of the animals in the study (e.g. cardiovascular and respiratory depression, CNS disturbance, hypothermia, reduction of food intake).

Examples

1) "Murine lymph node tumors arose in 11 of 12 mice that received N2-transduced human cells. The neo gene could be detected in murine cells as well as in human cells. Significant lymphoproliferation could be seen only in the murine pre-T cells. It took 5 months for murine leukemia to arise; the affected mice displayed symptoms of extreme sickness rapidly, with 5 of the 12 mice becoming moribund on exactly the same day (Figure 3), and 6 others becoming moribund within a 1-month period…Of the 12 mice that had received N2-transduced human cells, 11 had to be killed because they developed visibly enlarged lymph nodes and spleen, hunching, and decrease in body weight, as shown in Figure 3….The 12th mouse was observed carefully for 14 months; it did not show any signs of leukemia or other adverse events, and had no abnormal tissues when it was autopsied.…The mice were observed at least once daily for signs of illness, which were defined as any one or more of the following: weight loss, hunching, lethargy, rapid breathing, skin discoloration or irregularities, bloating, hemi-paresis, visibly enlarged lymph nodes, and visible solid tumors under the skin. Any signs of illness were logged as "adverse events" in the experiment, the mouse was immediately killed, and an autopsy was performed to establish the cause of illness." [200]

2) "Although procedures were based on those reported in the literature, dogs under Protocol 1 displayed high levels of stress and many experienced vomiting. This led us to significantly alter procedures in order to optimize the protocol for the purposes of our own fasting and postprandial metabolic studies." [201]

16c. Describe the humane endpoints established for the study and the frequency of monitoring.

Explanation

Humane endpoints are predetermined morphological, physiological and/or behavioural signs that define the circumstances under which an animal will be removed from an experimental study. The use of humane endpoints can help minimise harm while allowing the scientific objectives to be achieved [202]. Report the humane endpoints that were established for the specific study, species and strain. Include clear criteria of the clinical signs monitored [127], and clinical signs that led to euthanasia or other defined actions. Include details such as general welfare indicators (e.g. weight loss, reduced food intake, abnormal posture) and procedure-specific welfare indicators (e.g. tumour size in cancer studies [47], sensory motor deficits in stroke studies [203]).

Report the timing and frequency of monitoring, taking into consideration the normal circadian rhythm of the animal and timing of scientific procedures, as well as any increase in the frequency of monitoring (e.g. post-surgery recovery, critical times during disease studies, or following the observation of an adverse event). Publishing score sheets of the clinical signs that were monitored [204] can help guide other researchers to develop clinically relevant welfare assessments, particularly for studies reporting novel procedures.

Example

1) "Both the research team and the veterinary staff monitored animals twice daily. Health was monitored by weight (twice weekly), food and water intake, and general assessment of animal activity, panting, and fur condition…. The maximum size the tumors allowed to grow in the mice before euthanasia was 2000 mm^3." [205]

Item 17. Interpretation/scientific implications

17a. Interpret the results, taking into account the study objectives and hypotheses, current theory and other relevant studies in the literature.

Explanation

It is important to interpret the results of the study in the context of the study objectives (see **item 13 – Objectives**). For hypothesis-testing studies, interpretations should be restricted to the primary outcome (see i**tem 6 – Outcome measures**). Exploratory results derived from additional outcomes should not be described as conclusive, as they may be underpowered and less reliable.

Discuss the findings in the context of current theory, ideally with reference to a relevant systematic review, as individual studies do not provide a complete picture. If a systematic review is not available, take care to avoid selectively citing studies that corroborate the results, or only those that report statistically significant findings [206].

Where appropriate, describe any implications of the experimental methods or research findings for improving welfare standards or reducing the number of animals used in future studies (e.g. the use of a novel approach reduced the results' variability, thus enabling the use of smaller group sizes without losing statistical power). This may not be the primary focus of the research but reporting this information enables wider dissemination and uptake of refined techniques within the scientific community.

Examples

1) "This is in contrast to data provided by an 'intra-renal IL-18 overexpression' model [43], and may reflect an IL-18 concentration exceeding the physiologic range in the latter study." [207]
2) "The new apparatus shows potential for considerably reducing the number of animals used in memory tasks designed to detect potential amnesic properties of new drugs ... approximately 43,000 animals have been used in these tasks in the past 5 years but with the application of the continual trials apparatus we estimate that this could have been reduced to 26,000…. with the new paradigm the number of animals needed to obtain reliable results and maintain the statistical power of the tasks is greatly reduced." [208]
3) "In summary, our results show that IL-1Ra protects against brain injury and reduces neuroinflammation when administered peripherally to aged and comorbid animals at reperfusion or 3 hours later. These findings address concerns raised in a recent systematic review on IL-1Ra in stroke [209], and provide further supporting evidence for IL-1Ra as a lead candidate for the treatment of ischemic stroke." [210]

17b. Comment on the study limitations including potential sources of bias, limitations of the animal model, and imprecision associated with the results.

Explanation

Discussing the limitations of the work is important to place the findings in context, interpret the validity of the results, and ascribe a credibility level to its conclusions [211]. Limitations are unavoidable in scientific research, and describing them is essential to share experience, guide best practice, and aid the design of future experiments.

Discuss the quality of evidence presented in the study, and consider how appropriate the animal model is to the specific research question. A discussion on the rigour of the study design to isolate cause and effect (also known as internal validity [212]) should include whether potential risks of bias have been addressed [9].

Examples

1) "Although in this study we did not sample the source herds, the likelihood of these herds to be Influenza A virus (IAV) positive is high given the commonality of IAV infections in the Midwest. However, we cannot fully rule out the possibility that new gilts became infected with resident viruses after arrival to the herd. Although new gilts were placed into isolated designated areas and procedures were in place to minimize disease transmission (eg. isolation, vaccination), these areas or procedures might not have been able to fully contain infections within the designated areas." [213]
2) "Even though our data demonstrates that sustained systemic TLR9 stimulation aggravates diastolic HF in our model of gene-targeted diastolic HF, there are several limitations as to mechanistic explanations of causality, as well as extrapolations to clinical inflammatory disease states and other HF conditions. First, our pharmacological inflammatory model does not allow discrimination between effects caused by direct cardiac TLR9 stimulation to that of indirect effects mediated by systemic inflammation. Second, although several systemic inflammatory conditions have disturbances in the innate immune system as important features, and some of these again specifically encompassing distorted TLR9 signalling [34], sustained TLR9 stimulation does not necessarily represent a clinically relevant inflammatory condition. Finally, the cardiac myocyte SERCA2a KO model does not adequately represent the molecular basis for, or the clinical features of, diastolic HF." [214]

Item 18. Generalisability/translation

Comment on whether, and how, the findings of this study are likely to generalise to other species or experimental conditions, including any relevance to human biology (where appropriate).

Explanation

An important purpose of publishing research findings is to inform future research. In the context of animal studies, this might take the form of further *in vivo* research or another research domain (e.g. human clinical trial). Thoughtful consideration is warranted, as additional unnecessary animal studies are wasteful and unethical. Similarly, human clinical trials initiated based on insufficient or misleading animal research evidence increase research waste and negatively influence the risk-benefit balance for research participants [212, 215].

Consider whether the findings may be used to inform future research in a broadly similar context, or whether enough evidence has been accumulated in the literature to justify further research in another species or in humans. Discuss what (if any) further research may be required to allow generalisation or translation. Discuss and interpret the results in relation to current evidence, and in particular whether similar [216] or otherwise supportive [217] findings have been reported by other groups. Discuss the range of circumstances in which the effect is observed, and factors which may moderate that effect. Such factors could include for example the population (e.g. age, sex, strain, species), the intervention (e.g. different drugs of the same class), and the outcome measured (e.g. different approaches to assessing memory).

Examples

1) "Our results demonstrate that hDBS robustly modulates the mesolimbic network. This finding may hold clinical relevance for hippocampal DBS therapy in epilepsy cases, as connectivity in this network has previously been shown to be suppressed in mTLE. Further research is necessary to investigate potential DBS-induced restoration of MTLE-induced loss of functional connectivity in mesolimbic brain structures." [218]
2) "The tumor suppressor effects of GAS1 had been previously reported in cell cultures or in xenograft models, this is the first work in which the suppressor activity of murine Gas1 is reported for primary tumors in vivo. Recent advances in the design of safe vectors for transgene delivery may result in extrapolating our results to humans and so a promising field of research emerges in the area of hepatic, neoplastic diseases." [219]

Item 19. Protocol registration

Provide a statement indicating whether a protocol (including the research question, key design features, and analysis plan) was prepared before the study, and if and where this protocol was registered.

Explanation

Akin to the approach taken for clinical trials, protocol registration has emerged as a mechanism that is likely to improve the transparency of animal research [215, 220, 221]. Registering a protocol before the start of the experiment enables researchers to demonstrate that the hypothesis, approach and analysis were planned in advance and not shaped by data as they emerged; it enhances scientific rigour and protects the researcher against concerns about selective reporting of results [222, 223]. A protocol should consist of a) the question being addressed and the key features of the research that is proposed, such as the hypothesis being tested and the primary outcome measure (if applicable), the statistical analysis plan; and b) the laboratory procedures to be used to perform the planned experiment.

Protocols may be registered with different levels of completeness. For example, in the Registered Report format offered by an increasing number of journals, protocols undergo peer review and if accepted, the journal commits to publishing the completed research regardless of the results obtained [220].

Other online resources include the Open Science Framework [224], which is suitable to deposit PHISPS (Population; Hypothesis; Intervention; Statistical Analysis Plan; Primary; Outcome Measure; Sample Size Calculation) protocols [225] and provide researchers with the flexibility to embargo the preregistration and keep it from public view until the research is published, and selectively share it with reviewers and editors. The EDA can also be used to generate a time-stamped PDF, which sets out key elements of the experimental design [18]. This

can be used to demonstrate that the study conduct, analysis and reporting were not unduly driven by emerging data. As a minimum we recommend registering protocols containing all PHISPS components as outlined above.

Provide a statement indicating whether or not any protocol was prepared before the study, and if applicable, the location of its registration. Where there have been deviations from the protocol, describe the rationale for these changes in the publication so that readers can take this into account when assessing the findings.

Examples

1) "A detailed description of all protocols can be found in the Registered Report (Kandela et al., 2015). Additional detailed experimental notes, data, and analysis are available on the Open Science Framework (OSF) (RRID: SCR_003238) (https://osf.io/xu1g2/; Mantis et al., 2016)." [226]
2)"To maximise the objectivity of the presented analyses, we preregistered this study with its two hypotheses, its planned methods, and its complete plan of data analysis before the start of data collection (https://osf.io/fh8eq/), and we closely adhered to our plan… All statistical analyses closely followed our preregistered analysis plan (https://osf.io/fh8eq/)." [227]
3) "We preregistered our analyses with the Open Science Framework which facilitates reproducibility and open collaboration in science research (Foster & Deardorff, 2017). Our preregistration: Sheldon and Griffith (2017), was carried out to limit the number of analyses conducted and to validate our commitment to testing a limited number of a priori hypotheses. Our methods are consistent with this preregistration (Sheldon & Griffith, 2017)." [228]

Item 20. Data access

Provide a statement describing if and where study data are available.

Explanation

A data sharing statement describes how others can access the data on which the paper is based. Sharing adequately annotated data allows others to replicate data analyses, so that results can be independently tested and verified. Data sharing allows the data to be repurposed and new datasets to be created by combining data from multiple studies (e.g. to be used in secondary analyses). This allows others to explore new topics and increases the impact of the study, potentially preventing unnecessary use of animals and providing more value for money. Access to raw data also facilitates text and automated data mining [229].

An increasing number of publishers and funding bodies require authors or grant holders to make their data publicly available [230]. Journal articles with accompanying data may be cited more frequently [231]. Datasets can also be independently cited in their own right, which provides additional credit for authors. This practice is gaining increasing recognition and acceptance [232].

Where possible, make available all data that contribute to summary estimates or claims presented in the paper. Data should follow the FAIR guiding principles [233], that is data are findable, accessible (i.e. don't use outdated file types), interoperable (can be used on multiple platforms and with multiple software packages) and re-usable (i.e. have adequate data descriptors).

Data can be made publicly available via a structured, specialised (domain-specific), open access repository such as those maintained by NCBI (https://www.ncbi.nlm.nih.gov/) or EBI (https://www.ebi.ac.uk/). If such a repository is not available, data can be deposited in unstructured but publicly available repositories (e.g. Figshare (https://figshare.com/), Dryad (https://datadryad.org/), Zenodo (https://zenodo.org/) or Open Science Framework (https://osf.io/)). There are also search platforms to identify relevant repositories with rigorous standards (e.g. FairSharing (https://fairsharing.org/) and re3data (https://www.re3data.org/).

Examples

1) "Data Availability Statement: All data are available from Figshare at http://dx.doi.org/10.6084/m9.figshare.1288935." [234]

2) "A fundamental goal in generating this dataset is to facilitate access to spiny mouse transcript sequence information for external collaborators and researchers. The sequence reads and metadata are available from the NCBI (PRJNA342864) and assembled transcriptomes (Trinity_v2.3.2 and tr2aacds_v2) are available from the Zenodo repository (https://doi.org/10.5281/zenodo.808870), however accessing and utilizing this data can be challenging for researchers lacking bioinformatics expertise. To address this problem we are hosting a SequenceServer32 BLAST-search website (http://spinymouse.erc.monash.edu/sequenceserver/http://spinymouse.erc.monash.edu/sequenceserver/). This resource provides a user-friendly interface to access sequence information from the tr2aacds_v2 assembly (to explore annotated protein-coding transcripts) and/or the Trinity_v2.3.2 assembly (to explore non-coding transcripts)." [235]

Item 21. Declaration of interests

21a. Declare any potential conflicts of interest, including financial and non-financial. If none exist, this should be stated.

Explanation

A competing or conflict of interest is anything that interferes with (or could be perceived as interfering with) the full and objective presentation, analysis, and interpretation of the research. Competing or conflicts of interest can be financial or non-financial, professional or personal. They can exist in institutions, in teams, or with individuals. Potential competing interests are considered in peer review, editorial and publication decisions; the aim is to ensure transparency, and in most cases, a declaration of a conflict of interest does not obstruct the publication or review process.

Examples are provided in Box 8. If unsure, declare all potential conflicts, including both perceived and real conflicts of interest [236].

Box 8: Examples of competing or conflicts of interest

Financial:

Funding and other payments received or expected by the authors directly arising from the publication of the study, or funding or other payments from an organisation with an interest in the outcome of the work.

Non-financial:

Research that may benefit the individual or institution in terms of goods in kind. This includes unpaid advisory position in a government, non-government organisation or commercial organisations.

Affiliations:

Employed by, on the advisory board or a member of an organisation with an interest in the outcome of the work.

Intellectual property:

Patents or trademarks owned by someone or their organisation. This also includes the potential exploitation of the scientific advance being reported for the institution, the authors, or the research funders.

Personal:

Friends, family, relationships, and other close personal connections to people who may potentially benefit financially or in other ways from the research.

Ideology:

Beliefs or activism (e.g. political or religious) relevant to the work. Membership of a relevant advocacy or lobbying organisation.

Examples

1) "The study was funded by Gubra ApS. LSD; PJP ; GH ; KF and HBH are employed by Gubra ApS. JJ and NV are the owners of Gubra ApS. Gubra ApS provided support in the form of materials and salaries for authors LSD ; PJP ; GH ; KF ; HBH ; JJ and NV." [237]
2) "The authors have declared that no competing interests exist." [238].

21b. List all funding sources (including grant identifier) and the role of the funder(s) in the design, analysis and reporting of the study.

Explanation

The identification of funding sources allows the reader to assess any competing interests, and any potential sources of bias. For example, bias, as indicated by a prevalence of more favourable outcomes, has been demonstrated for clinical research funded by industry compared to studies funded by other sources [239-241]. Evidence for preclinical research also indicates that funding sources may influence the interpretation of study outcomes [236, 242].

Report the funding information including the financial supporting body(s) and any grant identifier(s). Include the role of the funder in the design, analysis, reporting and/or or decision to publish. If the research did not receive specific funding, but was performed as part of the employment of the authors, name the employer.

Examples

1) "Support was provided by the Italian Ministry of Health: Current research funds PRC 2010/001 [http://www.salute.gov.it/] to MG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." [243]
2) "This study was financially supported by the Tuberculosis and Lung Research Center of Tabriz University of Medical Sciences and the Research Council of University of Tabriz. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." [244]
3) "This work was supported by the salary paid to AEW. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." [245]

Acknowledgements

We would like to acknowledge the late Doug Altman's contribution to this project, Doug was a dedicated member of the working group and his input into the guidelines' revision has been invaluable.

Competing interests

AA: editor in chief of the British Journal of Pharmacology. WJB, ICC and ME: authors of the original ARRIVE guidelines. WJB: serves on the Independent Statistical Standing Committee of the funder CHDI foundation. AC, CJM, MMcL and ESS: involved in the IICARus trial. ME, MMcL and ESS: have received funding from NC3Rs. ME: sits on the MRC ERPIC panel. STH: chair of the NC3Rs board, trusteeship of the BLF, Kennedy Trust, DSRU and CRUK, member of Governing Board, Nuffield Council of Bioethics, member Science Panel for Health (EU H2020), founder and NEB Director Synairgen, consultant Novartis, Teva and AZ, chair MRC/GSK EMINENT Collaboration. VH, KL, EJP and NPdS: NC3Rs staff, role includes promoting the ARRIVE guidelines. CJMcC: shareholdings in Hindawi, on the publishing board of the Royal Society, on the EU Open Science policy platform. MMcL, NPdS, CJMcC, ESS, TS and HW: members of EQIPD. MMcL: member of the Animals in Science Committee. NPdS and TS: associate editors of BMJ Open Science. OP: vice president of Academia Europaea, senior executive editor of the Journal of Physiology, member of the Board of the European Commission's SAPEA (Science Advice for Policy

by European Academies). FR: NC3Rs board member, shareholdings in AstraZeneca and GSK. PR: member of the University of Florida Institutional Animal Care and Use Committee, Editorial board member of Shock. ESS: editor in chief of BMJ Open Science. SDS: role is to provide expertise and does not represent the opinion of the NIH. TS: shareholdings in Johnson & Johnson. SA, MTA, MB, UD, PG, DWH, NAK, and KR declared no conflict of interest.

References

1. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One. 2009;4(11):e7824. Epub 2009/12/04. doi: 10.1371/journal.pone.0007824. PubMed PMID: 19956596; PubMed Central PMCID: PMCPMC2779358.

2. Florez-Vargas O, Brass A, Karystianis G, Bramhall M, Stevens R, Cruickshank S, et al. Bias in the reporting of sex and age in biomedical research on mouse models. Elife. 2016;5. Epub 2016/03/05. doi: 10.7554/eLife.13615. PubMed PMID: 26939790; PubMed Central PMCID: PMCPMC4821800.

3. Weissgerber TL, Garcia-Valencia O, Garovic VD, Milic NM, Winham SJ. Why we need to report more than 'data were analyzed by t-tests or anova'. Elife. 2018;7. Epub 2018/12/24. doi: 10.7554/eLife.36163. PubMed PMID: 30574870; PubMed Central PMCID: PMCPMC6326723.

4. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of bias in reports of in vivo research: A focus for improvement. PLoS Biol. 2015;13(10):e1002273. Epub 2015/10/16. doi: 10.1371/journal.pbio.1002273. PubMed PMID: 26460723; PubMed Central PMCID: PMCPMC4603955.

5. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLoS Biol. 2010;8(6):e1000412. Epub 2010/07/09. doi: 10.1371/journal.pbio.1000412. PubMed PMID: 20613859; PubMed Central PMCID: PMCPMC2893951.

6. Baker D, Lidster K, Sottomayor A, Amor S. Two years later: Journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. PLoS Biol. 2014;12(1):e1001756. Epub 2014/01/11. doi: 10.1371/journal.pbio.1001756. PubMed PMID: 24409096; PubMed Central PMCID: PMCPMC3883646.

7. Gulin JE, Rocco DM, Garcia-Bournissen F. Quality of reporting and adherence to ARRIVE guidelines in animal studies for chagas disease preclinical drug research: A systematic review. PLoS Negl Trop Dis. 2015;9(11):e0004194. Epub 2015/11/21. doi: 10.1371/journal.pntd.0004194. PubMed PMID: 26587586; PubMed Central PMCID: PMCPMC4654562.

8. Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, et al. The devil is in the details: Incomplete reporting in preclinical animal research. PLoS ONE. 2016;11(11):e0166733. doi: 10.1371/journal.pone.0166733.

9. Reichlin TS, Vogt L, Würbel H. The researchers' view of scientific rigor—survey on the conduct and reporting of in vivo research. PLoS ONE. 2016;11(12):e0165999. doi: 10.1371/journal.pone.0165999.

10. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: Support for the ARRIVE (animal research: Reporting of in vivo experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. PLoS One. 2018;13(5):e0197882. Epub 2018/05/26. doi: 10.1371/journal.pone.0197882. PubMed PMID: 29795636; PubMed Central PMCID: PMCPMC5967836.

11. Hair K, Macleod MR, Sena ES, Sena ES, Hair K, Macleod MR, et al. A randomised controlled trial of an intervention to improve compliance with the ARRIVE guidelines (iicarus). Research Integrity and Peer Review. 2019;4(1):12. doi: 10.1186/s41073-019-0069-3.

12. McGrath JC, Lilley E. Implementing guidelines on reporting research using animals (ARRIVE etc.): New requirements for publication in bjp. Br J Pharmacol. 2015;172(13):3189-93. Epub 2015/05/13. doi: 10.1111/bph.12955. PubMed PMID: 25964986; PubMed Central PMCID: PMCPMC4500358.

13. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2019: Updated guidelines for reporting animal research. Submitted to Plos Biology. 2019.

14. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? PLOS Medicine. 2010;7(3):e1000245. doi: 10.1371/journal.pmed.1000245.

15. NC3Rs. Experimental design assistant version 1 [22/05/2019]. Available from: https://eda.nc3rs.org.uk/.

16. Ruxton G. Experimental design for the life sciences. New York, NY: Oxford University Press; 2016. pages cm p.

17. Festing MF, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR J. 2002;43(4):244-58. Epub 2002/10/23. PubMed PMID: 12391400.

18. Percie du Sert N, Bamsey I, Bate ST, Berdoy M, Clark RA, Cuthill I, et al. The experimental design assistant. PLoS Biol. 2017;15(9):e2003779. Epub 2017/09/29. doi: 10.1371/journal.pbio.2003779. PubMed PMID: 28957312; PubMed Central PMCID: PMCPMC5634641.

19. Nguyen TT, Bazzoli C, Mentre F. Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. Stat Med. 2012;31(11-12):1043-58. Epub 2011/10/04. doi: 10.1002/sim.4390. PubMed PMID: 21965170.

20. Hill D, Chen L, Snaar-Jagalska E, Chaudhry B. Embryonic zebrafish xenograft assay of human cancer metastasis [version 2; referees: 2 approved]. F1000Research. 2018;7(1682). doi: 10.12688/f1000research.16659.2.

21. Hurlbert SH. Pseudoreplication and the design of ecological field experiments. Ecological Monographs. 1984;54(2):187-211. doi: 10.2307/1942661.

22. Lazic SE. The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? BMC Neuroscience. 2010;11(1):5. doi: 10.1186/1471-2202-11-5.

23. Burdge GC, Lillycrop KA, Jackson AA, Gluckman PD, Hanson MA. The nature of the growth pattern and of the metabolic response to fasting in the rat are dependent upon the dietary protein and folic acid intakes of their pregnant dams and post-weaning fat consumption. Br J Nutr. 2008;99(3):540-9. Epub 2007/09/01. doi: 10.1017/S0007114507815819. PubMed PMID: 17761015; PubMed Central PMCID: PMCPMC2493056.

24. Bate ST, Clark RA. The design and statistical analysis of animal experiments. Cambridge, United Kingdom: Cambridge University Press; 2014. xv, 310 pages p.

25. Lazic SE, Clarke-Williams CJ, Munafò MR. What exactly is 'n' in cell culture and animal experiments? PLOS Biology. 2018;16(4):e2005282. doi: 10.1371/journal.pbio.2005282.

26. NC3Rs. Experimental unit 2015 [cited 2019 21 March]. Available from: https://eda.nc3rs.org.uk/experimental-design-unit.

27. Kwan S, King J, Grenier J, Yan J, Jiang X, Roberson M, et al. Maternal choline supplementation during normal murine pregnancy alters the placental epigenome: Results of an exploratory study. Nutrients. 2018;10(4):417. PubMed PMID: doi:10.3390/nu10040417.

28. Karp NA, Mason J, Beaudet AL, Benjamini Y, Bower L, Braun RE, et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. Nature Communications. 2017;8:15475. doi: 10.1038/ncomms15475.

29. Ribeiro FdAS, Vasquez LA, Fernandes JBK, Sakomura NK. Feeding level and frequency for freshwater angelfish. Revista Brasileira de Zootecnia. 2012;41:1550-4.

30. Grasselli G, Rossi S, Musella A, Gentile A, Loizzo S, Muzio L, et al. Abnormal nmda receptor function exacerbates experimental autoimmune encephalomyelitis. British journal of pharmacology. 2013;168(2):502-17. doi: 10.1111/j.1476-5381.2012.02178.x. PubMed PMID: 22924679.

31. Gray LE, Jr., Ostby J, Furr J, Price M, Veeramachaneni DN, Parks L. Perinatal exposure to the phthalates dehp, bbp, and dinp, but not dep, dmp, or dotp, alters sexual differentiation of the male rat. Toxicol Sci. 2000;58(2):350-65. Epub 2000/12/02. PubMed PMID: 11099647.

32. Vahidy F, Schäbitz W-R, Fisher M, Aronowski J. Reporting standards for preclinical studies of stroke therapy. Stroke. 2016;47(10):2435.

33. Muhlhausler BS, Bloomfield FH, Gillman MW. Whole animal experiments should be more like human randomized controlled trials. PLoS Biol. 2013;11(2):e1001481. doi: 10.1371/journal.pbio.1001481.

34. Jennions MD, Møller AP. A survey of the statistical power of research in behavioral ecology and animal behavior. Behavioral Ecology. 2003;14(3):438-45. doi: 10.1093/beheco/14.3.438.

35. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365-76. doi: 10.1038/nrn3475.

36. Würbel H. More than 3rs: The importance of scientific validity for harm-benefit analysis of animal research. Lab animal. 2017;46(4):164.

37. Walker M, Fureix C, Palme R, Newman JA, Ahloy Dallaire J, Mason G. Mixed-strain housing for female c57bl/6, dba/2, and balb/c mice: Validating a split-plot design that promotes refinement and reduction. BMC Medical Research Methodology. 2016;16(1):11. doi: 10.1186/s12874-016-0113-7.

38. Shaw R, Festing MF, Peers I, Furlong L. Use of factorial designs to optimize animal experiments and reduce animal use. ILAR J. 2002;43(4):223-32. Epub 2002/10/23. PubMed PMID: 12391398.

39. Neumann K, Grittner U, Piper SK, Rex A, Florez-Vargas O, Karystianis G, et al. Increasing efficiency of preclinical research by group sequential designs. PLoS Biol. 2017;15(3):e2001307. Epub 2017/03/11. doi: 10.1371/journal.pbio.2001307. PubMed PMID: 28282371; PubMed Central PMCID: PMCPMC5345756.

40. Faul F, Erdfelder E, Lang AG, Buchner A. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175-91. Epub 2007/08/19. PubMed PMID: 17695343.

41. Festing MFW. On determining sample size in experiments involving laboratory animals. Laboratory Animals. 2018;52(4):341-50. doi: 10.1177/0023677217738268.

42. Reynolds PS, McCarter J, Sweeney C, Mohammed BM, Brophy DF, Fisher B, et al. Informing efficient pilot development of animal trauma models through quality improvement strategies. Lab Anim. 2018:23677218802999. Epub 2018/10/10. doi: 10.1177/0023677218802999. PubMed PMID: 30296892.

43. Bate S. How to decide your sample size when the power calculation is not straightforward NC3Rs.org.uk: NC3Rs; 2018 [updated 01/08/2018; cited 2018 02/08/2018]. Available from: https://www.nc3rs.org.uk/news/how-decide-your-sample-size-when-power-calculation-not-straightforward.

44. Bustamante R, Daza MA, Canfrán S, García P, Suárez M, Trobo I, et al. Comparison of the postoperative analgesic effects of cimicoxib, buprenorphine and their combination in healthy dogs undergoing ovariohysterectomy. Veterinary Anaesthesia and Analgesia. 2018;45(4):545-56. doi: 10.1016/j.vaa.2018.01.003.

45. Spin JR, Oliveira GJPLd, Spin-Neto R, Pires JR, Tavares HS, Ykeda F, et al. Avaliação histomorfométrica da associação entre biovidro e osso bovino liofilizado no tratamento de defeitos ósseos críticos criados em calvárias de ratos. Estudo piloto. Revista de Odontologia da UNESP. 2015;44:37-43.

46. Salkind NJ. Encyclopedia of research design: Sage; 2010.

47. Workman P, Aboagye EO, Balkwill F, Balmain A, Bruder G, Chaplin DJ, et al. Guidelines for the welfare and use of animals in cancer research. British Journal Of Cancer. 2010;102:1555. doi: 10.1038/sj.bjc.6605642.

48. Sena ES, Jeffreys AL, Cox SF, Sastra SA, Churilov L, Rewell S, et al. The benefit of hypothermia in experimental ischemic stroke is not affected by pethidine. International journal of stroke : official journal of the International Stroke Society. 2013;8(3):180-5. doi: 10.1111/j.1747-4949.2012.00834.x. PubMed PMID: 22759525.

49. Kafkafi N, Agassi J, Chesler EJ, Crabbe JC, Crusio WE, Eilam D, et al. Reproducibility and replicability of rodent phenotyping in preclinical studies. Neuroscience & Biobehavioral Reviews. 2018;87:218-32. doi: 10.1016/j.neubiorev.2018.01.003.

50. Scott S, Kranz JE, Cole J, Lincecum JM, Thompson K, Kelly N, et al. Design, power, and interpretation of studies in the standard murine model of als. Amyotrophic Lateral Sclerosis. 2008;9(1):4-15. doi: 10.1080/17482960701856300.

51. Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, et al. Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. PLoS Biol. 2016;14(1):e1002331. Epub 2016/01/05. doi: 10.1371/journal.pbio.1002331. PubMed PMID: 26726833; PubMed Central PMCID: PMCPMC4699644.

52. Rice ASC, Morland R, Huang W, Currie GL, Sena ES, Macleod MR. Transparency in the reporting of in vivo pre-clinical pain research: The relevance and implications of the ARRIVE (animal research: Reporting in vivo experiments) guidelines. Scandinavian Journal of Pain. 2013;4(2):58-62. doi: 10.1016/j.sjpain.2013.02.002.

53. Genther-Schroeder ON, Branine ME, Hansen SL. Effects of increasing supplemental dietary zn concentration on growth performance and carcass characteristics in finishing steers fed ractopamine hydrochloride. J Anim Sci. 2018;96(5):1903-13. Epub 2018/05/08. doi: 10.1093/jas/sky094. PubMed PMID: 29733414.

54. Castiglioni L, Colazzo F, Fontana L, Colombo GI, Piacentini L, Bono E, et al. Evaluation of left ventricle function by regional fractional area change (rfac) in a mouse model of myocardial infarction secondary to valsartan treatment. PLOS ONE. 2015;10(8):e0135778. doi: 10.1371/journal.pone.0135778.

55. Brent LJ, Heilbronner SR, Horvath JE, Gonzalez-Martinez J, Ruiz-Lambides A, Robinson AG, et al. Genetic origins of social networks in rhesus macaques. Sci Rep. 2013;3:1042. Epub 2013/01/11. doi: 10.1038/srep01042. PubMed PMID: 23304433; PubMed Central PMCID: PMCPMC3540398.

56. Gouveia K, Hurst JL. Optimising reliability of mouse performance in behavioural testing: The major role of non-aversive handling. Sci Rep. 2017;7:44999. Epub 2017/03/23. doi: 10.1038/srep44999. PubMed PMID: 28322308; PubMed Central PMCID: PMC5359560.

57. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995;273(5):408-12. Epub 1995/02/01. PubMed PMID: 7823387.

58. Schulz KF, Grimes DA. Allocation concealment in randomised trials: Defending against deciphering. Lancet. 2002;359(9306):614-8. Epub 2002/02/28. doi: 10.1016/s0140-6736(02)07750-4. PubMed PMID: 11867132.

59. Chalmers TC, Celano P, Sacks HS, Smith Jr H. Bias in treatment assignment in controlled clinical trials. New England Journal of Medicine. 1983;309(22):1358-61.

60. Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, et al. The need for randomization in animal trials: An overview of systematic reviews. PLoS ONE. 2014;9(6):e98856. doi: 10.1371/journal.pone.0098856.

61. Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. Improving the translational hit of experimental treatments in multiple sclerosis. Multiple Sclerosis Journal. 2010;16(9):1044-55. doi: doi:10.1177/1352458510379612. PubMed PMID: 20685763.

62. Taves DR. Minimization: A new method of assigning patients to treatment and control groups. Clin Pharmacol Ther. 1974;15(5):443-53. Epub 1974/05/01. PubMed PMID: 4597226.

63. Saint-Mont U. Randomization does not help much, comparability does. PLOS ONE. 2015;10(7):e0132102. doi: 10.1371/journal.pone.0132102.

64. Laajala TD, Jumppanen M, Huhtaniemi R, Fey V, Kaur A, Knuuttila M, et al. Optimized design and analysis of preclinical intervention studies in vivo. Scientific Reports. 2016;6:30723. doi: 10.1038/srep30723.

65. Zhao S, Kang R, Deng T, Luo L, Wang J, Li E, et al. Comparison of two cannulation methods for assessment of intracavernosal pressure in a rat model. PLoS One. 2018;13(2):e0193543. Epub 2018/02/28. doi: 10.1371/journal.pone.0193543. PubMed PMID: 29486011; PubMed Central PMCID: PMCPMC5828359.

66. Jansen of Lorkeers SJ, Gho JMIH, Koudstaal S, van Hout GPJ, Zwetsloot PPM, van Oorschot JWM, et al. Xenotransplantation of human cardiomyocyte progenitor cells does not improve cardiac function in a porcine model of chronic ischemic heart failure. Results from a randomized, blinded, placebo controlled trial. PLOS ONE. 2015;10(12):e0143953. doi: 10.1371/journal.pone.0143953.

67. Ishida A, Mutoh T, Ueyama T, Bando H, Masubuchi S, Nakahara D, et al. Light activates the adrenal gland: Timing of gene expression and glucocorticoid release. Cell Metab. 2005;2(5):297-307. Epub 2005/11/08. doi: 10.1016/j.cmet.2005.09.009. PubMed PMID: 16271530.

68. Kang M, Ragan BG, Park JH. Issues in outcomes research: An overview of randomization techniques for clinical trials. J Athl Train. 2008;43(2):215-21. Epub 2008/03/18. doi: 10.4085/1062-6050-43.2.215. PubMed PMID: 18345348; PubMed Central PMCID: PMCPMC2267325.

69. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: Why randomise? BMJ. 1999;318(7192):1209. Epub 1999/04/30. PubMed PMID: 10221955; PubMed Central PMCID: PMCPMC1115595.

70. Altman DG, Bland JM. Treatment allocation by minimisation. BMJ. 2005;330(7495):843. Epub 2005/04/09. doi: 10.1136/bmj.330.7495.843. PubMed PMID: 15817555; PubMed Central PMCID: PMCPMC556084.

71. El-Agroudy NN, El-Naga RN, El-Razeq RA, El-Demerdash E. Forskolin, a hedgehog signalling inhibitor, attenuates carbon tetrachloride-induced liver fibrosis in rats. British Journal of Pharmacology. 2016;173(22):3248-60. doi: 10.1111/bph.13611.

72. Carrillo M, Migliorati F, Bruls R, Han Y, Heinemans M, Pruis I, et al. Repeated witnessing of conspecifics in pain: Effects on emotional contagion. PLOS ONE. 2015;10(9):e0136979. doi: 10.1371/journal.pone.0136979.

73. Del Bianco Benedeti P, Paulino PVR, Marcondes MI, Maciel IFS, da Silva MC, Faciola AP. Partial replacement of ground corn with glycerol in beef cattle diets: Intake, digestibility, performance, and carcass characteristics. PLOS ONE. 2016;11(1):e0148224. doi: 10.1371/journal.pone.0148224.

74. Nuzzo R. How scientists fool themselves–and how they can stop. Nature News. 2015;526(7572):182.

75. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: Systematic review of trials with both blinded and non-blinded outcome assessors. BMJ. 2012;344. doi: 10.1136/bmj.e1119.

76. Rosenthal R, Fode KL. The effect of experimenter bias on the performance of the albino rat. Behavioral Science. 1963;8(3):183-9. doi: 10.1002/bs.3830080302.

77. Rosenthal R, Lawson R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. Journal of Psychiatric Research. 1964;2(2):61-72. doi: 10.1016/0022-3956(64)90003-2.

78. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of nxy-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke. 2008;39(10):2824-9. Epub 2008/07/19. doi: 10.1161/strokeaha.108.515957. PubMed PMID: 18635842.

79. Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. Improving the translational hit of experimental treatments in multiple sclerosis. Mult Scler. 2010;16(9):1044-55. Epub 2010/08/06. doi: 1352458510379612 [pii] 10.1177/1352458510379612. PubMed PMID: 20685763.

80. Gómez de Segura IA, Bustamante R, Daza MA, Canfrán S, García P, Suárez M, et al. Comparison of the postoperative analgesic effects of cimicoxib, buprenorphine and their combination in healthy dogs undergoing ovariohysterectomy. Veterinary Anaesthesia and Analgesia. 2018. doi: 10.1016/j.vaa.2018.01.003.

81. Neumann A-M, Abele J, Kirschstein T, Engelmann R, Sellmann T, Köhling R, et al. Mycophenolate mofetil prevents the delayed t cell response after pilocarpine-induced status epilepticus in mice. PLOS ONE. 2017;12(11):e0187330. doi: 10.1371/journal.pone.0187330.

82. Hsieh LS, Wen JH, Miyares L, Lombroso PJ, Bordey A. Outbred cd1 mice are as suitable as inbred c57bl/6j mice in performing social tasks. Neuroscience Letters. 2017;637:142-7. doi: 10.1016/j.neulet.2016.11.035.

83. Claudino MA, Leiria LOS, da Silva FH, Alexandre EC, Renno A, Mónica FZ, et al. Urinary bladder dysfunction in transgenic sickle cell disease mice. PLOS ONE. 2015;10(8):e0133996. doi: 10.1371/journal.pone.0133996.

84. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological Science. 2012;23(5):524-32. doi: 10.1177/0956797611430953.

85. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012;490(7419):187-91. Epub 2012/10/13. doi: 10.1038/nature11556. PubMed PMID: 23060188; PubMed Central PMCID: PMCPMC3511845.

86. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLOS Biology. 2015;13(3):e1002106. doi: 10.1371/journal.pbio.1002106.

87. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. Nature Human Behaviour. 2017;1:0021. doi: 10.1038/s41562-016-0021.

88. Emmrich J, Neher J, Boehm-Sturm P, Endres M, Dirnagl U, Harms C. Stage 1 registered report: Effect of deficient phagocytosis on neuronal survival and neurological outcome after temporary middle cerebral artery occlusion (tmcao) [version 3; referees: 2 approved]. F1000Research. 2018;6(1827). doi: 10.12688/f1000research.12537.3.

89. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: The "statistical analyses and methods in the published literature" or the sampl guidelines. Int J Nurs Stud. 2015;52(1):5-9. Epub 2014/12/03. doi: 10.1016/j.ijnurstu.2014.09.006. PubMed PMID: 25441757.

90. Polo J, Rodríguez C, Ródenas J, Russell LE, Campbell JM, Crenshaw JD, et al. Ultraviolet light (uv) inactivation of porcine parvovirus in liquid plasma and effect of uv irradiated spray dried porcine plasma on performance of weaned pigs. PLOS ONE. 2015;10(7):e0133008. doi: 10.1371/journal.pone.0133008.

91. Garner JP, Dufour B, Gregg LE, Weisker SM, Mench JA. Social and husbandry factors affecting the prevalence and severity of barbering ('whisker trimming') by laboratory mice. Applied Animal Behaviour Science. 2004;89(3-4):263-82. doi: 10.1016/j.applanim.2004.07.004. PubMed PMID: WOS:000225168300007.

92. Podrini C, Cambridge EL, Lelliott CJ, Carragher DM, Estabel J, Gerdin A-K, et al. High-fat feeding rapidly induces obesity and lipid derangements in c57bl/6n mice. Mammalian genome : official journal of the International Mammalian Genome Society. 2013;24(5-6):240-51. doi: 10.1007/s00335-013-9456-0. PubMed PMID: 23712496.

93. Ohishi T, Wang L, Akane H, Shiraki A, Itahashi M, Mitsumori K, et al. Transient suppression of late-stage neuronal progenitor cell differentiation in the hippocampal dentate gyrus of rat offspring after maternal exposure to nicotine. Arch Toxicol. 2014;88(2):443-54. Epub 2013/07/31. doi: 10.1007/s00204-013-1100-y. PubMed PMID: 23892646.

94. Ruxton G, Colegrave N. Experimental design for the life sciences. Fourth ed: Oxford University Press; 2017.

95. Nemeth M, Millesi E, Wagner K-H, Wallner B. Sex-specific effects of diets high in unsaturated fatty acids on spatial learning and memory in guinea pigs. PLOS ONE. 2015;10(10):e0140485. doi: 10.1371/journal.pone.0140485.

96. Keck VA, Edgerton DS, Hajizadeh S, Swift LL, Dupont WD, Lawrence C, et al. Effects of habitat complexity on pair-housed zebrafish. J Am Assoc Lab Anim Sci. 2015;54(4):378-83. Epub 2015/08/01. PubMed PMID: 26224437; PubMed Central PMCID: PMCPMC4521571.

97. Clayton JA, Collins FS. Policy: Nih to balance sex in cell and animal studies. Nature News. 2014;509(7500):282.

98. Shapira S, Sapir M, Wengier A, Grauer E, Kadar T. Aging has a complex effect on a rat model of ischemic stroke. Brain research. 2002;925(2):148-58. doi: 10.1016/s0006-8993(01)03270-x. PubMed PMID: 11792363.

99. Vital M, Harkema JR, Rizzo M, Tiedje J, Brandenberger C. Alterations of the murine gut microbiome with age and allergic airway disease. J Immunol Res. 2015;2015:892568. Epub 2015/06/20. doi: 10.1155/2015/892568. PubMed PMID: 26090504; PubMed Central PMCID: PMCPMC4451525.

100. Bouwknecht JA, Paylor R. Behavioral and physiological mouse assays for anxiety: A survey in nine mouse strains. Behav Brain Res. 2002;136(2):489-501. Epub 2002/11/14. PubMed PMID: 12429412.

101. Simon MM, Greenaway S, White JK, Fuchs H, Gailus-Durner V, Wells S, et al. A comparative phenotypic and genomic analysis of c57bl/6j and c57bl/6n mouse strains. Genome Biol. 2013;14(7):R82. Epub 2013/08/02. doi: 10.1186/gb-2013-14-7-r82. PubMed PMID: 23902802; PubMed Central PMCID: PMCPMC4053787.

102. Jackson SJ, Andrews N, Ball D, Bellantuono I, Gray J, Hachoumi L, et al. Does age matter? The impact of rodent age on study outcomes. Laboratory Animals. 2017;51(2):160-9. doi: 10.1177/0023677216653984. PubMed PMID: 27307423.

103. Khokha MK, Chung C, Bustamante EL, Gaw LW, Trott KA, Yeh J, et al. Techniques and probes for the study of xenopus tropicalis development. Dev Dyn. 2002;225(4):499-510. Epub 2002/11/28. doi: 10.1002/dvdy.10184. PubMed PMID: 12454926.

104. Okva K, Nevalainen T, Pokk P. The effect of cage shelf on the behaviour of male c57bl/6 and balb/c mice in the elevated plus maze test. Lab Anim. 2013;47(3):220-2. Epub 2013/06/14. doi: 10.1177/0023677213489280. PubMed PMID: 23760964.

105. Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, Francoijs KJ, Stunnenberg HG, et al. A hierarchy of h3k4me3 and h3k27me3 acquisition in spatial gene regulation in xenopus embryos. Dev Cell. 2009;17(3):425-34. Epub 2009/09/18. doi: 10.1016/j.devcel.2009.08.005. PubMed PMID: 19758566; PubMed Central PMCID: PMCPMC2746918.

106. Mahler Convenor M, Berard M, Feinstein R, Gallagher A, Illgen-Wilcke B, Pritchett-Corning K, et al. Felasa recommendations for the health monitoring of mouse, rat, hamster, guinea pig and rabbit colonies in breeding and experimental units. Lab Anim. 2014;48(3):178-92. Epub 2014/02/06. doi: 10.1177/0023677213516312. PubMed PMID: 24496575.

107. Baker DG. Natural pathogens of laboratory mice, rats, and rabbits and their effects on research. Clinical Microbiology Reviews. 1998;11(2):231-66. doi: 10.1128/cmr.11.2.231.

108. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, et al. Endogenous enterobacteriaceae underlie variation in susceptibility to salmonella infection. Nat Microbiol. 2019. Epub 2019/03/27. doi: 10.1038/s41564-019-0407-8. PubMed PMID: 30911125.

109. Holmdahl R, Malissen B. The need for littermate controls. Eur J Immunol. 2012;42(1):45-7. Epub 2012/01/04. doi: 10.1002/eji.201142048. PubMed PMID: 22213045.

110. Mallapaty S. In the name of reproducibility. Lab Animal. 2018;47(7):178-81. doi: 10.1038/s41684-018-0095-7.

111. Sundberg JP, Schofield PN. Commentary: Mouse genetic nomenclature:Standardization of strain, gene, and protein symbols. Veterinary Pathology. 2010;47(6):1100-4. doi: 10.1177/0300985810374837. PubMed PMID: 20685919.

112. Montoliu L, Whitelaw CBA. Using standard nomenclature to adequately name transgenes, knockout gene alleles and any mutation associated to a genetically modified mouse strain. Transgenic Research. 2011;20(2):435-40. doi: 10.1007/s11248-010-9428-z.

113. Huang W, Feng Y, Liang J, Yu H, Wang C, Wang B, et al. Loss of microrna-128 promotes cardiomyocyte proliferation and heart regeneration. Nat Commun. 2018;9(1):700. Epub 2018/02/18. doi: 10.1038/s41467-018-03019-z. PubMed PMID: 29453456; PubMed Central PMCID: PMCPMC5816015.

114. Ranson A, Cheetham CEJ, Fox K, Sengpiel F. Homeostatic plasticity mechanisms are required for juvenile, but not adult, ocular dominance plasticity. Proceedings of the National Academy of Sciences. 2012;109(4):1311-6. doi: 10.1073/pnas.1112204109.

115. Clarkson JM, Dwyer DM, Flecknell PA, Leach MC, Rowe C. Handling method alters the hedonic value of reward in laboratory mice. Sci Rep. 2018;8(1):2448. Epub 2018/02/07. doi: 10.1038/s41598-018-20716-3. PubMed PMID: 29402923; PubMed Central PMCID: PMCPMC5799408.

116. Hurst JL, West RS. Taming anxiety in laboratory mice. Nat Methods. 2010;7(10):825-6. Epub 2010/09/14. doi: 10.1038/nmeth.1500. PubMed PMID: 20835246.

117. Hewitt JA, Brown LL, Murphy SJ, Grieder F, Silberberg SD. Accelerating biomedical discoveries through rigor and transparency. ILAR J. 2017;58(1):115-28. Epub 2017/06/03. doi: 10.1093/ilar/ilx011. PubMed PMID: 28575443; PubMed Central PMCID: PMCPMC6279133.

118. Almeida JL, Cole KD, Plant AL. Standards for cell line authentication and beyond. PLoS Biol. 2016;14(6):e1002476. Epub 2016/06/15. doi: 10.1371/journal.pbio.1002476. PubMed PMID: 27300367; PubMed Central PMCID: PMCPMC4907466.

119. Leary SL, Underwood W, Anthony R, Cartner S, Corey D, Grandin T, et al. Avma guidelines for the euthanasia of animals: 2013 edition. 2013.

120. Bandrowski AE, Martone ME. Rrids: A simple step toward improving reproducibility through rigor and transparency of experimental methods. Neuron. 2016;90(3):434-6. Epub 2016/05/07. doi: 10.1016/j.neuron.2016.04.030. PubMed PMID: 27151636; PubMed Central PMCID: PMCPMC5854161.

121. Bandrowski A, Brush M, Grethe JS, Haendel MA, Kennedy DN, Hill S, et al. The resource identification initiative: A cultural shift in publishing. J Comp Neurol. 2016;524(1):8-22. Epub 2015/11/26. doi: 10.1002/cne.23913. PubMed PMID: 26599696; PubMed Central PMCID: PMCPMC4684178.

122. Teytelman L, Stoliartchouk A. Protocols.Io: Reducing the knowledge that perishes because we do not publish it. Information Services & Use. 2016;35(1/2):109-15. PubMed PMID: 109101228.

123. Reynolds PS, Fisher BJ, McCarter J, Sweeney C, Martin EJ, Middleton P, et al. Interventional vitamin c: A strategy for attenuation of coagulopathy and inflammation in a swine polytrauma model. J Trauma Acute Care Surg. 2018. Epub 2018/03/15. doi: 10.1097/ta.0000000000001844. PubMed PMID: 29538225.

124. Bauters D, Bedossa P, Lijnen HR, Hemmeryckx B. Functional role of adamts5 in adiposity and metabolic health. PLoS One. 2018;13(1):e0190595. Epub 2018/01/03. doi: 10.1371/journal.pone.0190595. PubMed PMID: 29293679; PubMed Central PMCID: PMCPMC5749841.

125. Bartlang MS, Neumann ID, Slattery DA, Uschold-Schmidt N, Kraus D, Helfrich-Förster C, et al. Time matters: Pathological effects of repeated psychosocial stress during the active, but not inactive, phase of male mice. Journal of Endocrinology. 2012;215(3):425-37. doi: 10.1530/joe-12-0267.

126. Paul AK, Gueven N, Dietis N. Morphine dosing strategy plays a key role in the generation and duration of the produced antinociceptive tolerance. Neuropharmacology. 2017;121:158-66. Epub 2017/04/30. doi: 10.1016/j.neuropharm.2017.04.034. PubMed PMID: 28450061.

127. Hawkins P, Morton DB, Burman O, Dennison N, Honess P, Jennings M, et al. A guide to defining and implementing protocols for the welfare assessment of laboratory animals: Eleventh report of the bvaawf/frame/rspca/ufaw joint working group on refinement. Laboratory Animals. 2011;45(1):1-13. doi: 10.1258/la.2010.010031.

128. Hagemo JS, Jørgensen JJ, Ostrowski SR, Holtan A, Gundersen Y, Johansson PI, et al. Changes in fibrinogen availability and utilization in an animal model of traumatic coagulopathy. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine. 2013;21(1):56. doi: 10.1186/1757-7241-21-56.

129. Emery M, Nanchen N, Preitner F, Ibberson M, Roduit R. Biological characterization of gene response to insulin-induced hypoglycemia in mouse retina. PLOS ONE. 2016;11(2):e0150266. doi: 10.1371/journal.pone.0150266.

130. Holmes AM, Emmans CJ, Coleman R, Smith TE, Hosie CA. Effects of transportation, transport medium and re-housing on xenopus laevis (daudin). Gen Comp Endocrinol. 2018;266:21-8. Epub 2018/03/17. doi: 10.1016/j.ygcen.2018.03.015. PubMed PMID: 29545087.

131. Conour LA, Murray KA, Brown MJ. Preparation of animals for research--issues to consider for rodents and rabbits. ILAR J. 2006;47(4):283-93. Epub 2006/09/12. PubMed PMID: 16963809.

132. Obernier JA, Baldwin RL. Establishing an appropriate period of acclimatization following transportation of laboratory animals. ILAR Journal. 2006;47(4):364-9. doi: 10.1093/ilar.47.4.364.

133. Krahn DD, Gosnell BA, Majchrzak MJ. The anorectic effects of crh and restraint stress decrease with repeated exposures. Biol Psychiatry. 1990;27(10):1094-102. Epub 1990/05/15. PubMed PMID: 2340320.

134. Pitman DL, Ottenweller JE, Natelson BH. Plasma corticosterone levels during repeated presentation of two intensities of restraint stress: Chronic stress and habituation. Physiol Behav. 1988;43(1):47-55. Epub 1988/01/01. PubMed PMID: 3413250.

135. Brock AJ, Goody SMG, Mead AN, Sudwarts A, Parker MO, Brennan CH. Assessing the value of the zebrafish conditioned place preference model for predicting human abuse potential. J Pharmacol Exp Ther. 2017;363(1):66-79. Epub 2017/08/10. doi: 10.1124/jpet.117.242628. PubMed PMID: 28790193; PubMed Central PMCID: PMCPMC5602714.

136. Turner PV, Brabb T, Pekow C, Vasbinder MA. Administration of substances to laboratory animals: Routes of administration and factors to consider. Journal of the American Association for Laboratory Animal Science : JAALAS. 2011;50(5):600-13. PubMed PMID: PMC3189662.

137. Fueger BJ, Czernin J, Hildebrandt I, Tran C, Halpern BS, Stout D, et al. Impact of animal handling on the results of 18f-fdg pet studies in mice. J Nucl Med. 2006;47(6):999-1006. Epub 2006/06/03. PubMed PMID: 16741310.

138. Aslan Y, Tadjuidje E, Zorn AM, Cha SW. High-efficiency non-mosaic crispr-mediated knock-in and indel mutation in f0 xenopus. Development. 2017;144(15):2852-8. Epub 2017/07/12. doi: 10.1242/dev.152967. PubMed PMID: 28694259; PubMed Central PMCID: PMCPMC5560047.

139. Elmorsy S, Soliman G, Rashed L, Elgendy H. The response to sedative doses of propofol and dexmedetomidine in a prenatal valproate autistic rat model. Kasr Al Ainy Medical Journal. 2018;24(1):32-9. doi: 10.4103/kamj.kamj_37_17.

140. Ash JA, Lu H, Taxier LR, Long JM, Yang Y, Stein EA, et al. Functional connectivity with the retrosplenial cortex predicts cognitive aging in rats. Proc Natl Acad Sci U S A. 2016;113(43):12286-91. Epub 2016/10/30. doi: 10.1073/pnas.1525309113. PubMed PMID: 27791017; PubMed Central PMCID: PMCPMC5087009.

141. Altman DG. Why we need confidence intervals. World J Surg. 2005;29(5):554-6. Epub 2005/04/14. doi: 10.1007/s00268-005-7911-0. PubMed PMID: 15827844.

142. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340. doi: 10.1136/bmj.c869.

143. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: A practical guide for biologists. Biol Rev Camb Philos Soc. 2007;82(4):591-605. Epub 2007/10/20. doi: 10.1111/j.1469-185X.2007.00027.x. PubMed PMID: 17944619.

144. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. Ann Intern Med. 1990;113(1):69-76. Epub 1990/07/01. PubMed PMID: 2190518.

145. Berwanger O, Ribeiro RA, Finkelsztejn A, Watanabe M, Suzumura EA, Duncan BB, et al. The quality of reporting of trial abstracts is suboptimal: Survey of major general medical journals. J Clin Epidemiol. 2009;62(4):387-92. Epub 2008/11/18. doi: 10.1016/j.jclinepi.2008.05.013. PubMed PMID: 19010643.

146. Can OS, Yilmaz AA, Hasdogan M, Alkaya F, Turhan SC, Can MF, et al. Has the quality of abstracts for randomised controlled trials improved since the release of consolidated standards of reporting trial guideline for abstract reporting? A survey of four high-profile anaesthesia journals. Eur J Anaesthesiol. 2011;28(7):485-92. Epub 2010/11/03. doi: 10.1097/EJA.0b013e32833fb96f. PubMed PMID: 21037480.

147. Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. JAMA. 1999;281(12):1110-1. Epub 1999/04/03. PubMed PMID: 10188662.

148. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: The spiin randomized controlled trial. J Clin Oncol. 2014;32(36):4120-6. Epub 2014/11/19. doi: 10.1200/JCO.2014.56.7503. PubMed PMID: 25403215.

149. Toledo AC, Sakoda CPP, Perini A, Pinheiro NM, Magalhães RM, Grecco S, et al. Flavonone treatment reverses airway inflammation and remodelling in an asthma murine model. British Journal of Pharmacology. 2013;168(7):1736-49. doi: 10.1111/bph.12062. PubMed PMID: 23170811.

150. DeBoer SP, Garner JP, McCain RR, Lay Jr DC, Eicher SD, Marchant-Forde JN. An initial investigation into the effects of isolation and enrichment on the welfare of laboratory pigs housed in the pigturn® system, assessed using tear staining, behaviour, physiology and haematology. Animal Welfare. 2015;24(1):15-27. doi: 10.7120/09627286.24.1.015.

151. Avey MT, Fenwick N, Griffin G. The use of systematic reviews and reporting guidelines to advance the implementation of the 3rs. Journal of the American Association for Laboratory Animal Science : JAALAS. 2015;54(2):153-62. PubMed PMID: PMC4382619.

152. Hooijmans CR, Ritskes-Hoitinga M. Progress in using systematic reviews of animal studies to improve translational research. PLOS Medicine. 2013;10(7):e1001482. doi: 10.1371/journal.pmed.1001482.

153. Donner DG, Elliott GE, Beck BR, Bulmer AC, Du Toit EF. Impact of diet-induced obesity and testosterone deficiency on the cardiovascular system: A novel rodent model representative of males with testosterone-deficient metabolic syndrome (tdmets). PLOS ONE. 2015;10(9):e0138019. doi: 10.1371/journal.pone.0138019.

154. Willner P. Validation criteria for animal models of human mental disorders: Learned helplessness as a paradigm case. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 1986;10(6):677-90. doi: 10.1016/0278-5846(86)90051-5.

155. Hunt RF, Girskis KM, Rubenstein JL, Alvarez–Buylla A, Baraban SC. Gaba progenitors grafted into the adult epileptic brain control seizures and abnormal behavior. Nature neuroscience. 2013;16(6):692-7. doi: 10.1038/nn.3392. PubMed PMID: PMC3665733.

156. Hamilton N, Sabroe I, Renshaw SA. A method for transplantation of human hscs into zebrafish, to replace humanised murine transplantation models. F1000Res. 2018;7:594. Epub 2019/01/05. doi: 10.12688/f1000research.14507.2. PubMed PMID: 29946444; PubMed Central PMCID: PMCPMC6008850.2.

157. Kimmelman J, Mogil JS, Dirnagl U. Distinguishing between exploratory and confirmatory preclinical research will improve translation. PLoS Biol. 2014;12(5):e1001863. doi: 10.1371/journal.pbio.1001863.

158. Camp DM, Loeffler DA, Farrah DM, Borneman JN, LeWitt PA. Cellular immune response to intrastriatally implanted allogeneic bone marrow stromal cells in a rat model of parkinson's disease. J Neuroinflammation. 2009;6:17. Epub 2009/06/09. doi: 10.1186/1742-2094-6-17. PubMed PMID: 19500379; PubMed Central PMCID: PMCPMC2700085.

159. Falk H, Forde PF, Bay ML, Mangalanathan UM, Hojman P, Soden DM, et al. Calcium electroporation induces tumor eradication, long-lasting immunity and cytokine responses in the ct26 colon cancer mouse model. OncoImmunology. 2017;6(5):e1301332. doi: 10.1080/2162402X.2017.1301332.

160. Bayne K, Turner PV. Animal welfare standards and international collaborations. ILAR J. 2019. Epub 2019/01/10. doi: 10.1093/ilar/ily024. PubMed PMID: 30624646.

161. Redfern WS, Tse K, Grant C, Keerie A, Simpson DJ, Pedersen JC, et al. Automated recording of home cage activity and temperature of individual rats housed in social groups: The rodent big brother project. PLoS ONE. 2017;12(9):e0181068. doi: 10.1371/journal.pone.0181068.

162. Wang X, Xue Q, Yan F, Liu J, Li S, Hu S. Ulinastatin protects against acute kidney injury in infant piglets model undergoing surgery on hypothermic low-flow cardiopulmonary bypass. PLOS ONE. 2015;10(12):e0144516. doi: 10.1371/journal.pone.0144516.

163. Berghof TV, van der Klein SA, Arts JA, Parmentier HK, van der Poel JJ, Bovenhuis H. Genetic and non-genetic inheritance of natural antibodies binding keyhole limpet hemocyanin in a purebred layer chicken line. PLoS One. 2015;10(6):e0131088. Epub 2015/06/27. doi: 10.1371/journal.pone.0131088. PubMed PMID: 26114750; PubMed Central PMCID: PMCPMC4482680.

164. Bailoo JD, Murphy E, Varholick JA, Novak J, Palme R, Würbel H. Evaluation of the effects of space allowance on measures of animal welfare in laboratory mice. Scientific Reports. 2018;8(1):713. doi: 10.1038/s41598-017-18493-6.

165. Kallnik M, Elvert R, Ehrhardt N, Kissling D, Mahabir E, Welzl G, et al. Impact of ivc housing on emotionality and fear learning in male c3heb/fej and c57bl/6j mice. Mamm Genome. 2007;18(3):173-86. Epub 2007/04/14. doi: 10.1007/s00335-007-9002-z. PubMed PMID: 17431719.

166. Holmes AM, Emmans CJ, Jones N, Coleman R, Smith TE, Hosie CA. Impact of tank background on the welfare of the african clawed frog, xenopus laevis (daudin). Applied Animal Behaviour Science. 2016;185:131-6. doi: 10.1016/j.applanim.2016.09.005. PubMed PMID: WOS:000390504700018.

167. Haseman JK, Ney E, Nyska A, Rao GN. Effect of diet and animal care/housing protocols on body weight, survival, tumor incidences, and nephropathy severity of f344 rats in chronic studies. Toxicol Pathol. 2003;31(6):674-81. Epub 2003/10/31. doi: 10.1080/01926230390241927. PubMed PMID: 14585736.

168. Morris JK, Bomhoff GL, Stanford JA, Geiger PC. Neurodegeneration in an animal model of parkinson's disease is exacerbated by a high-fat diet. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 2010;299(4):R1082-R90. doi: 10.1152/ajpregu.00449.2010.

169. Bayol SA, Farrington SJ, Stickland NC. A maternal 'junk food' diet in pregnancy and lactation promotes an exacerbated taste for 'junk food' and a greater propensity for obesity in rat offspring. Br J Nutr. 2007;98(4):843-51. Epub 2007/08/19. doi: 10.1017/S0007114507812037. PubMed PMID: 17697422.

170. Gaskill BN, Garner JP. Stressed out: Providing laboratory animals with behavioral control to reduce the physiological effects of stress. Lab Animal. 2017;46:142. doi: 10.1038/laban.1218.

171. Robinson I, Dowdall T, Meert TF. Development of neuropathic pain is affected by bedding texture in two models of peripheral nerve injury in rats. Neurosci Lett. 2004;368(1):107-11. Epub 2004/09/03. doi: 10.1016/j.neulet.2004.06.078. PubMed PMID: 15342144.

172. Kokolus KM, Capitano ML, Lee CT, Eng JW, Waight JD, Hylander BL, et al. Baseline tumor growth and immune control in laboratory mice are significantly influenced by subthermoneutral housing temperature. Proc Natl Acad Sci U S A. 2013;110(50):20176-81. Epub 2013/11/20. doi: 10.1073/pnas.1304291110. PubMed PMID: 24248371; PubMed Central PMCID: PMCPMC3864348.

173. Duke JL, Zammit TG, Lawson DM. The effects of routine cage-changing on cardiovascular and behavioral parameters in male sprague-dawley rats. Contemp Top Lab Anim Sci. 2001;40(1):17-20. Epub 2001/04/13. PubMed PMID: 11300670.

174. Prager E, Bergstrom H, Grunberg N, Johnson L. The importance of reporting housing and husbandry in rat research. Front Behav Neurosci. 2011;5(38). doi: 10.3389/fnbeh.2011.00038.

175. Rosenbaum MD, VandeWoude S, Johnson TE. Effects of cage-change frequency and bedding volume on mice and their microenvironment. Journal of the American Association for Laboratory Animal Science : JAALAS. 2009;48(6):763-73. PubMed PMID: PMC2786931.

176. Kappel S, Hawkins P, Mendl MT. To group or not to group? Good practice for housing male laboratory mice. Animals : an Open Access Journal from MDPI. 2017;7(12):88. doi: 10.3390/ani7120088. PubMed PMID: PMC5742782.

177. Van Loo PLP, Mol JA, Koolhaas JM, Van Zutphen BFM, Baumans V. Modulation of aggression in male mice: Influence of group size and cage size. Physiology & Behavior. 2001;72(5):675-83. doi: 10.1016/S0031-9384(01)00425-5.

178. Bleich A, Hansen AK. Time to include the gut microbiota in the hygienic standardisation of laboratory rodents. Comp Immunol Microbiol Infect Dis. 2012;35(2):81-92. Epub 2012/01/20. doi: 10.1016/j.cimid.2011.12.006. PubMed PMID: 22257867.

179. Dauchy RT, Dupepe LM, Ooms TG, Dauchy EM, Hill CR, Mao L, et al. Eliminating animal facility light-at-night contamination and its effect on circadian regulation of rodent physiology, tumor growth, and metabolism: A challenge in the relocation of a cancer research laboratory. J Am Assoc Lab Anim Sci. 2011;50(3):326-36. Epub 2011/06/07. PubMed PMID: 21640027; PubMed Central PMCID: PMCPMC3103282.

180. Chapillon P, Manneché C, Belzung C, Caston J. Rearing environmental enrichment in two inbred strains of mice: 1. Effects on emotional reactivity. Behavior Genetics. 1999;29(1):41-6. doi: 10.1023/a:1021437905913.

181. Hendershott TR, Cronin ME, Langella S, McGuinness PS, Basu AC. Effects of environmental enrichment on anxiety-like behavior, sociability, sensory gating, and spatial learning in male and female c57bl/6j mice. Behavioural Brain Research. 2016;314:215-25. doi: 10.1016/j.bbr.2016.08.004.

182. Garner JP. Stereotypies and other abnormal repetitive behaviors: Potential impact on validity, reliability, and replicability of scientific outcomes. Ilar j. 2005;46(2):106-17. Epub 2005/03/19. PubMed PMID: 15775020.

183. Gross AN-M, Engel AKJ, Würbel H. Simply a nest? Effects of different enrichments on stereotypic and anxiety-related behaviour in mice. Applied Animal Behaviour Science. 2011;134(3):239-45. doi: 10.1016/j.applanim.2011.06.020.

184. Wurbel H. Ideal homes? Housing effects on rodent brain and behaviour. Trends Neurosci. 2001;24(4):207-11. Epub 2001/03/16. PubMed PMID: 11250003.

185. Auvergne R, Déan C, El Bahh B, Arthaud S, lespinet-najib v, Rougier A, et al. Delayed kindling epileptogenesis and increased neurogenesis in adult rats housed in an enriched environment2002. 277-85 p.

186. Salvarrey-Strati A, Watson L, Blanchet T, Lu N, Glasson SS. The influence of enrichment devices on development of osteoarthritis in a surgically induced murine model. Ilar j. 2008;49(4):23-30. Epub 2008/10/14. PubMed PMID: 18849588.

187. Hannan AJ. Environmental enrichment and brain repair: Harnessing the therapeutic effects of cognitive stimulation and physical activity to enhance experience-dependent plasticity. Neuropathol Appl Neurobiol. 2014;40(1):13-25. Epub 2013/12/21. doi: 10.1111/nan.12102. PubMed PMID: 24354721.

188. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. Nature Methods. 2014;11:629. doi: 10.1038/nmeth.2935

189. Kotloski RJ, Sutula TP. Environmental enrichment: Evidence for an unexpected therapeutic influence. Exp Neurol. 2015;264:121-6. Epub 2014/12/09. doi: 10.1016/j.expneurol.2014.11.012. PubMed PMID: 25483395.

190. Jensen TL, Kiersgaard MK, Sorensen DB, Mikkelsen LF. Fasting of mice: A review. Lab Anim. 2013;47(4):225-40. Epub 2013/09/13. doi: 10.1177/0023677213501659. PubMed PMID: 24025567.

191. National Research Council. Guide for the care and use of laboratory animals. 8th ed: The National Academies Press; 2011.

192. European Commission. Directive 2010/63/eu of the european parliament and of the council of 22 september 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union [Internet]. 2010; 53. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32010L0063.

193. Heykants M, Mahabir E. Estrous cycle staging before mating led to increased efficiency in the production of pseudopregnant recipients without negatively affecting embryo transfer in mice. Theriogenology. 2016;85(5):813-21. Epub 2015/11/29. doi: 10.1016/j.theriogenology.2015.10.027. PubMed PMID: 26613855.

194. Gallastegui A, Cheung J, Southard T, Hume KR. Volumetric and linear measurements of lung tumor burden from non-gated micro-ct imaging correlate with histological analysis in a genetically engineered mouse model of non-small cell lung cancer. Lab Anim. 2018:23677218756457. Epub 2018/02/14. doi: 10.1177/0023677218756457. PubMed PMID: 29436921.

195. Jirkof P. Side effects of pain and analgesia in animal experimentation. Lab Anim (NY). 2017;46(4):123-8. Epub 2017/03/23. doi: 10.1038/laban.1216. PubMed PMID: 28328895.

196. Carbone L, Austin J. Pain and laboratory animals: Publication practices for better data reproducibility and better animal welfare. PLoS One. 2016;11(5):e0155001. Epub 2016/05/14. doi: 10.1371/journal.pone.0155001. PubMed PMID: 27171143; PubMed Central PMCID: PMCPMC4865140.

197. Gaspani L, Bianchi M, Limiroli E, Panerai AE, Sacerdote P. The analgesic drug tramadol prevents the effect of surgery on natural killer cell activity and metastatic colonization in rats. J Neuroimmunol. 2002;129(1-2):18-24. Epub 2002/08/06. PubMed PMID: 12161016.

198. Getty CM, Dilger RN. Moderate perinatal choline deficiency elicits altered physiology and metabolomic profiles in the piglet. PLoS One. 2015;10(7):e0133500. Epub 2015/07/22. doi: 10.1371/journal.pone.0133500. PubMed PMID: 26196148; PubMed Central PMCID: PMCPMC4510435.

199. Teilmann AC, Falkenberg MK, Hau J, Abelson KS. Comparison of silicone and polyurethane catheters for the catheterization of small vessels in mice. Lab Anim (NY). 2014;43(11):397-403. Epub 2014/10/22. doi: 10.1038/laban.570. PubMed PMID: 25333592.

200. Bauer G, Dao MA, Case SS, Meyerrose T, Wirthlin L, Zhou P, et al. In vivo biosafety model to assess the risk of adverse events from retroviral and lentiviral vectors. Mol Ther. 2008;16(7):1308-15. Epub 2008/05/08. doi: 10.1038/mt.2008.93. PubMed PMID: 18461052; PubMed Central PMCID: PMCPMC3013368.

201. Bellanger S, Benrezzak O, Battista MC, Naimi F, Labbe SM, Frisch F, et al. Experimental dog model for assessment of fasting and postprandial fatty acid metabolism: Pitfalls and feasibility. Lab Anim. 2015;49(3):228-40. Epub 2015/01/08. doi: 10.1177/0023677214566021. PubMed PMID: 25563731.

202. Hendriksen C, Morton D, Cussler K. Use of humane endpoints to minimise suffering. Howard B, T N, Peratta G, editors: CRC Press, Florida, USA; 2011.

203. Percie du Sert N, Alfieri A, Allan SM, Carswell HV, Deuchar GA, Farr TD, et al. The improve guidelines (ischaemia models: Procedural refinements of in vivo experiments). J Cereb Blood Flow Metab. 2017;37(11):3488-517. Epub 2017/08/12. doi: 10.1177/0271678X17709185. PubMed PMID: 28797196; PubMed Central PMCID: PMCPMC5669349.

204. Morton DB. A systematic approach for establishing humane endpoints. ILAR Journal. 2000;41(2):80-6. doi: 10.1093/ilar.41.2.80.

205. Muscella A, Vetrugno C, Cossa LG, Antonaci G, De Nuccio F, De Pascali SA, et al. In vitro and in vivo antitumor activity of [pt(o,o'-acac)(gamma-acac)(dms)] in malignant pleural mesothelioma. PLoS One. 2016;11(11):e0165154. Epub 2016/11/03. doi: 10.1371/journal.pone.0165154. PubMed PMID: 27806086; PubMed Central PMCID: PMCPMC5091852.

206. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383(9913):267-76. Epub 2014/01/15. doi: 10.1016/s0140-6736(13)62228-x. PubMed PMID: 24411647.

207. Schirmer B, Wedekind D, Glage S, Neumann D. Deletion of il-18 expression ameliorates spontaneous kidney failure in mrllpr mice. PLOS ONE. 2015;10(10):e0140173. doi: 10.1371/journal.pone.0140173.

208. Ameen-Ali KE, Eacott MJ, Easton A. A new behavioural apparatus to reduce animal numbers in multiple types of spontaneous object recognition paradigms in rats. Journal of neuroscience methods. 2012;211(1):66-76. doi: 10.1016/j.jneumeth.2012.08.006. PubMed PMID: 22917958.

209. Banwell V, Sena ES, Macleod MR. Systematic review and stratified meta-analysis of the efficacy of interleukin-1 receptor antagonist in animal models of stroke. Journal of Stroke and Cerebrovascular Diseases. 2009;18(4):269-76. doi: 10.1016/j.jstrokecerebrovasdis.2008.11.009.

210. Pradillo JM, Denes A, Greenhalgh AD, Boutin H, Drake C, McColl BW, et al. Delayed administration of interleukin-1 receptor antagonist reduces ischemic brain damage and inflammation in comorbid rats. Journal of Cerebral Blood Flow & Metabolism. 2012;32(9):1810-9. doi: 10.1038/jcbfm.2012.101.

211. Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. J Clin Epidemiol. 2007;60(4):324-9. Epub 2007/03/10. doi: 10.1016/j.jclinepi.2006.09.011. PubMed PMID: 17346604.

212. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy studies in investigator brochures: Do they enable risk–benefit assessment? PLOS Biology. 2018;16(4):e2004879. doi: 10.1371/journal.pbio.2004879.

213. Diaz A, Perez A, Sreevatsan S, Davies P, Culhane M, Torremorell M. Association between influenza a virus infection and pigs subpopulations in endemically infected breeding herds. PLOS ONE. 2015;10(6):e0129213. doi: 10.1371/journal.pone.0129213.

214. Dhondup Y, Sjaastad I, Scott H, Sandanger Ø, Zhang L, Haugstad SB, et al. Sustained toll-like receptor 9 activation promotes systemic and cardiac inflammation, and aggravates diastolic heart failure in serca2a ko mice. PLOS ONE. 2015;10(10):e0139715. doi: 10.1371/journal.pone.0139715.

215. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. The Lancet. 2014;383(9912):156-65. doi: 10.1016/S0140-6736(13)62229-1.

216. Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. PLOS Biology. 2018;16(2):e2003693. doi: 10.1371/journal.pbio.2003693.

217. Munafò MR, Davey Smith G. Robust research needs many lines of evidence. Nature. 2018;553(7689):399-401. Epub 2018/01/26. doi: 10.1038/d41586-018-01023-3. PubMed PMID: 29368721.

218. Van Den Berge N, Vanhove C, Descamps B, Dauwe I, van Mierlo P, Vonck K, et al. Functional mri during hippocampal deep brain stimulation in the healthy rat brain. PLOS ONE. 2015;10(7):e0133245. doi: 10.1371/journal.pone.0133245.

219. Sacilotto N, Castillo J, Riffo-Campos ÁL, Flores JM, Hibbitt O, Wade-Martins R, et al. Growth arrest specific 1 (gas1) gene overexpression in liver reduces the in vivo progression of murine hepatocellular carcinoma and partially restores gene expression levels. PLOS ONE. 2015;10(7):e0132477. doi: 10.1371/journal.pone.0132477.

220. Chambers CD, Feredoes E, Muthukumaraswamy SD, Etchells PJ. Instead of "playing the game" it is time to change the rules: Registered reports at *aims neuroscience* and beyond. AIMS Neuroscience. 2014;1(1):4-17. doi: DOI: 10.3934/Neuroscience2014.1.4.

221. Nosek BA, Lakens D. Registered reports a method to increase the credibility of published results. Social Psychology. 2014;45(3):137-41. doi: 10.1027/1864-9335/a000192. PubMed PMID: WOS:000336836900001.

222. Kaplan RM, Irvin VL. Likelihood of null effects of large nhlbi clinical trials has increased over time. Plos One. 2015;10(8). doi: ARTN e0132382 10.1371/journal.pone.0132382. PubMed PMID: WOS:000359061400022.

223. Allen C, Mehler D. Open science challenges, benefits and tips in early career and beyond. PsyArXiv. 2018. doi: 10.31234/osf.io/3czyt.

224. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci U S A. 2018;115(11):2600-6. Epub 2018/03/14. doi: 10.1073/pnas.1708274114. PubMed PMID: 29531091; PubMed Central PMCID: PMCPMC5856500.

225. Macleod M, Howells D. Protocols for laboratory research. Evidence-based Preclinical Medicine. 2016;3(2):e00021. doi: doi:10.1002/ebm2.21.

226. Mantis C, Kandela I, Aird F, Reproducibility Project: Cancer B. Replication study: Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. eLife. 2017;6:e17584. doi: 10.7554/eLife.17584.

227. Jeronimo S, Khadraoui M, Wang DP, Martin K, Lesku JA, Robert KA, et al. Plumage color manipulation has no effect on social dominance or fitness in zebra finches. Behavioral Ecology. 2018;29(2):459-67. doi: 10.1093/beheco/arx195. PubMed PMID: WOS:000427885600027.

228. Sheldon EL, Griffith SC. Embryonic heart rate predicts prenatal development rate, but is not related to post-natal growth rate or activity level in the zebra finch (taeniopygia guttata). Ethology. 2018;124(11):829-37. doi: 10.1111/eth.12817. PubMed PMID: WOS:000449485800006.

229. Kafkafi N, Mayo CL, Elmer GI. Mining mouse behavior for patterns predicting psychiatric drug classification. Psychopharmacology (Berl). 2014;231(1):231-42. Epub 2013/08/21. doi: 10.1007/s00213-013-3230-6. PubMed PMID: 23958942.

230. Stodden V, Guo P, Ma Z. Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. PLOS ONE. 2013;8(6):e67111. doi: 10.1371/journal.pone.0067111.

231. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. PLoS ONE. 2007;2(3):e308. doi: 10.1371/journal.pone.0000308.

232. DataCitationSynthesisGroup. Joint declaration of data citation principles San Diego CA: FORCE11; 2014 [cited 2018 22 May].

233. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The fair guiding principles for scientific data management and stewardship. Scientific Data. 2016;3:160018. doi: 10.1038/sdata.2016.18.

234. Federer LM, Lu Y-L, Joubert DJ, Welsh J, Brandys B. Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. PLoS ONE. 2015;10(6):e0129506. doi: 10.1371/journal.pone.0129506.

235. Mamrot J, Legaie R, Ellery SJ, Wilson T, Seemann T, Powell DR, et al. De novo transcriptome assembly for the spiny mouse (acomys cahirinus). Scientific Reports. 2017;7(1):8996. doi: 10.1038/s41598-017-09334-7.

236. Bero L, Anglemyer A, Vesterinen H, Krauth D. The relationship between study sponsorship, risks of bias, and research outcomes in atrazine exposure studies conducted in non-human animals: Systematic review and meta-analysis. Environment International. 2016;92-93:597-604. doi: 10.1016/j.envint.2015.10.011.

237. Dalbøge LS, Pedersen PJ, Hansen G, Fabricius K, Hansen HB, Jelsing J, et al. A hamster model of diet-induced obesity for preclinical evaluation of anti-obesity, anti-diabetic and lipid modulating agents. PLOS ONE. 2015;10(8):e0135634. doi: 10.1371/journal.pone.0135634.

238. Garcia de la serrana D, Vieira VLA, Andree KB, Darias M, Estévez A, Gisbert E, et al. Development temperature has persistent effects on muscle growth responses in gilthead sea bream. PLOS ONE. 2012;7(12):e51884. doi: 10.1371/journal.pone.0051884.

239. Lundh A, Sismondo S, Lexchin J, Busuioc OA, Bero L. Industry sponsorship and research outcome. Cochrane Database of Systematic Reviews. 2012;(12). doi: 10.1002/14651858.MR000033.pub2. PubMed PMID: MR000033.

240. Popelut A, Valet F, Fromentin O, Thomas A, Bouchard P. Relationship between sponsorship and failure rate of dental implants: A systematic approach. PLOS ONE. 2010;5(4):e10274. doi: 10.1371/journal.pone.0010274.

241. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. Bmj. 2003;326(7400):1167-70. Epub 2003/05/31. doi: 10.1136/bmj.326.7400.1167. PubMed PMID: 12775614; PubMed Central PMCID: PMCPMC156458.

242. Krauth D, Anglemyer A, Philipps R, Bero L. Nonindustry-sponsored preclinical studies on statins yield greater efficacy estimates than industry-sponsored studies: A meta-analysis. PLoS Biology. 2014;12(1):e1001770. doi: 10.1371/journal.pbio.1001770.

243. Genchi M, Prati P, Vicari N, Manfredini A, Sacchi L, Clementi E, et al. Francisella tularensis: No evidence for transovarial transmission in the tularemia tick vectors dermacentor reticulatus and ixodes ricinus. PLOS ONE. 2015;10(8):e0133593. doi: 10.1371/journal.pone.0133593.

244. Kolahian S, Sadri H, Shahbazfar AA, Amani M, Mazadeh A, Mirani M. The effects of leucine, zinc, and chromium supplements on inflammatory events of the respiratory system in type 2 diabetic rats. PLOS ONE. 2015;10(7):e0133374. doi: 10.1371/journal.pone.0133374.

245. Eyre-Walker A, Stoletzki N. The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. PLOS Biology. 2013;11(10):e1001675. doi: 10.1371/journal.pbio.1001675.

RESEARCH METHODS & REPORTING

CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials

Kenneth F Schulz,[1] Douglas G Altman,[2] David Moher,[3] for the CONSORT Group

EDITORIAL by Antes
RESEARCH, p 697

[1]Family Health International, Research Triangle Park, NC 27709, USA
[2]Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford
[3]Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada
Correspondence to: K F Schulz
kschulz@fhi.org

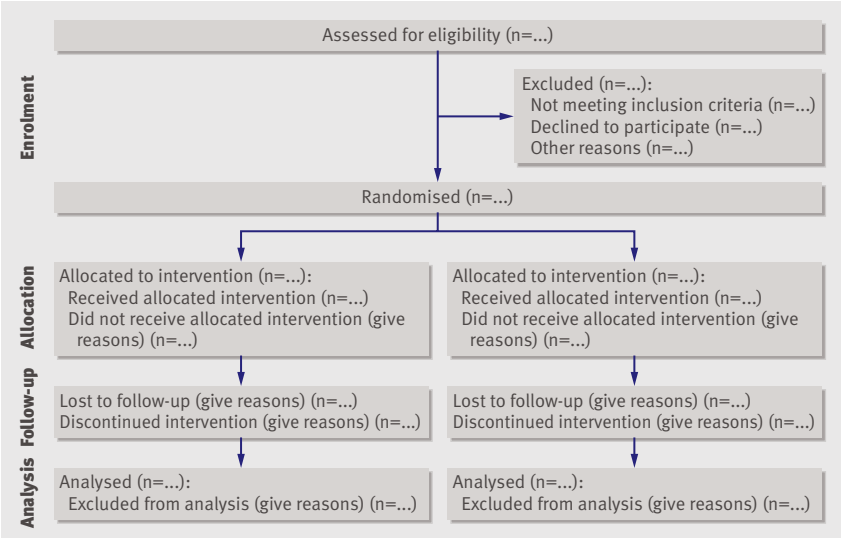
Accepted: 9 December 2009

Cite this as: *BMJ* 2010;340:c332
doi: 10.1136/bmj.c332

The CONSORT statement is used worldwide to improve the reporting of randomised controlled trials. **Kenneth Schulz and colleagues** describe the latest version, CONSORT 2010, which updates the reporting guideline based on new methodological evidence and accumulating experience

Randomised controlled trials, when appropriately designed, conducted, and reported, represent the gold standard in evaluating healthcare interventions. However, randomised trials can yield biased results if they lack methodological rigour.[1] To assess a trial accurately, readers of a published report need complete, clear, and transparent information on its methodology and findings. Unfortunately, attempted assessments frequently fail because authors of many trial reports neglect to provide lucid and complete descriptions of that critical information.[2-4]

That lack of adequate reporting fuelled the development of the original CONSORT (Consolidated Standards of Reporting Trials) statement in 1996[5] and its revision five years later.[6-8] While those statements improved the reporting quality for some randomised controlled trials,[9 10] many trial reports still remain inadequate.[2] Furthermore, new methodological evidence and additional experience has accumulated since the last revision in 2001. Consequently, we organised a CONSORT Group meeting to update the 2001 statement.[6-8] We introduce here the result of that process, CONSORT 2010.



Flow diagram of the progress through the phases of a parallel randomised trial of two groups (that is, enrolment, intervention allocation, follow-up, and data analysis)

Intent of CONSORT 2010

The CONSORT 2010 Statement is this paper including the 25 item checklist in the table and the flow diagram. It provides guidance for reporting all randomised controlled trials, but focuses on the most common design type—individually randomised, two group, parallel trials. Other trial designs, such as cluster randomised trials and non-inferiority trials, require varying amounts of additional information. CONSORT extensions for these designs,[11 12] and other CONSORT products, can be found through the CONSORT website (www.consort-statement.org). Along with the CONSORT statement, we have updated the explanation and elaboration article,[13] which explains the inclusion of each checklist item, provides methodological background, and gives published examples of transparent reporting.

Diligent adherence by authors to the checklist items facilitates clarity, completeness, and transparency of reporting. Explicit descriptions, not ambiguity or omission, best serve the interests of all readers. Note that the CONSORT 2010 Statement does not include recommendations for designing, conducting, and analysing trials. It solely addresses the reporting of what was done and what was found.

Nevertheless, CONSORT does indirectly affect design and conduct. Transparent reporting reveals deficiencies in research if they exist. Thus, investigators who conduct inadequate trials, but who must transparently report, should not be able to pass through the publication process without revelation of their trial's inadequacies. That emerging reality should provide impetus to improved trial design and conduct in the future, a secondary indirect goal of our work. Moreover, CONSORT can help researchers in designing their trial.

Background to CONSORT

Efforts to improve the reporting of randomised controlled trials accelerated in the mid-1990s, spurred partly by methodological research. Researchers had shown for many years that authors reported such trials poorly, and empirical evidence began to accumulate that some poorly conducted or poorly reported aspects of trials were associated with bias.[14] Two initiatives aimed at developing reporting guidelines culminated in one of us (DM) and Drummond Rennie organising the first CONSORT statement in 1996.[5]

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item
Title and abstract		
	1a	Identification as a randomised trial in the title
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts[21 31])
Introduction		
Background and objectives	2a	Scientific background and explanation of rationale
	2b	Specific objectives or hypotheses
Methods		
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants	4a	Eligibility criteria for participants
	4b	Settings and locations where the data were collected
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
	6b	Any changes to trial outcomes after the trial commenced, with reasons
Sample size	7a	How sample size was determined
	7b	When applicable, explanation of any interim analyses and stopping guidelines
Randomisation:		
Sequence generation	8a	Method used to generate the random allocation sequence
	8b	Type of randomisation; details of any restriction (such as blocking and block size)
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how
	11b	If relevant, description of the similarity of interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses
Results		
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome
	13b	For each group, losses and exclusions after randomisation, together with reasons
Recruitment	14a	Dates defining the periods of recruitment and follow-up
	14b	Why the trial ended or was stopped
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[28])
Discussion		
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses
Generalisability	21	Generalisability (external validity, applicability) of the trial findings
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
Other information		
Registration	23	Registration number and name of trial registry
Protocol	24	Where the full trial protocol can be accessed, if available
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration[13] for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials,[11] non-inferiority and equivalence trials,[12] non-pharmacological treatments,[32] herbal interventions,[33] and pragmatic trials.[34] Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Further methodologi cal research on similar topics reinforced earlier findings[15] and fed into the revision of 2001.[6-8] Subsequently, the expanding body of methodological research informed the refinement of CONSORT 2010. More than 700 studies comprise the CONSORT database (located on the CONSORT website), which provides the empirical evidence to underpin the CONSORT initiative.

Indeed, CONSORT Group members continually monitor the literature. Information gleaned from these efforts provides an evidence base on which to update the CONSORT statement. We add, drop, or modify items based on that evidence and the recommendations of the CONSORT Group, an international and eclectic group of clinical trialists, statisticians, epidemiologists, and biomedical editors. The CONSORT Executive (KFS, DGA, DM) strives for a balance of established and emerging researchers. The membership of the group is dynamic. As our work expands in response to emerging projects and needed expertise, we invite new members to contribute. As such, CONSORT continually assimilates new ideas and perspectives. That process informs the continually evolving CONSORT statement.

Over time, CONSORT has garnered much support. More than 400 journals, published around the world and in many languages, have explicitly supported the CONSORT

statement. Many other healthcare journals support it without our knowledge. Moreover, thousands more have implicitly supported it with the endorsement of the CONSORT statement by the International Committee of Medical Journal Editors (www.icmje.org). Other prominent editorial groups, the Council of Science Editors and the World Association of Medical Editors, officially support CONSORT. That support seems warranted: when used by authors and journals, CONSORT seems to improve reporting.[9]

Development of CONSORT 2010

Thirty one members of the CONSORT 2010 Group met in Montebello, Canada, in January 2007 to update the 2001 CONSORT statement. In addition to the accumulating evidence relating to existing checklist items, several new issues had come to prominence since 2001. Some participants were given primary responsibility for aggregating and synthesising the relevant evidence on a particular checklist item of interest. Based on that evidence, the group deliberated the value of each item. As in prior CONSORT versions, we kept only those items deemed absolutely fundamental to reporting a randomised controlled trial. Moreover, an item may be fundamental to a trial but not included, such as approval by an institutional ethical review board, because funding bodies strictly enforce ethical review and medical journals usually address reporting ethical review in their instructions for authors. Other items may seem desirable, such as reporting on whether on-site monitoring was done, but a lack of empirical evidence or any consensus on their value cautions against inclusion at this point. The CONSORT 2010 Statement thus addresses the minimum criteria, although that should not deter authors from including other information if they consider it important.

After the meeting, the CONSORT Executive convened teleconferences and meetings to revise the checklist. After seven major iterations, a revised checklist was distributed to the larger group for feedback. With that feedback, the executive met twice in person to consider all the comments and to produce a penultimate version. That served as the basis for writing the first draft of this paper, which was then distributed to the group for feedback. After consideration of their comments, the executive finalised the statement.

The CONSORT Executive then drafted an updated explanation and elaboration manuscript, with assistance from other members of the larger group. The substance of the 2007 CONSORT meeting provided the material for the update. The updated explanation and elaboration manuscript was distributed to the entire group for additions, deletions, and changes. That final iterative process converged to the CONSORT 2010 Explanation and Elaboration.[13]

Box 1 | Noteworthy general changes in CONSORT 2010 Statement

- We simplified and clarified the wording, such as in items 1, 8, 10, 13, 15, 16, 18, 19, and 21
- We improved consistency of style across the items by removing the imperative verbs that were in the 2001 version
- We enhanced specificity of appraisal by breaking some items into sub-items. Many journals expect authors to complete a CONSORT checklist indicating where in the manuscript the items have been addressed. Experience with the checklist noted pragmatic difficulties when an item comprised multiple elements. For example, item 4 addresses eligibility of participants and the settings and locations of data collection. With the 2001 version, an author could provide a page number for that item on the checklist, but might have reported only eligibility in the paper, for example, and not reported the settings and locations. CONSORT 2010 relieves obfuscations and forces authors to provide page numbers in the checklist for both eligibility and settings

Changes in CONSORT 2010

The revision process resulted in evolutionary, not revolutionary, changes to the checklist (table), and the flow diagram was not modified except for one word (figure). Moreover, because other reporting guidelines augmenting the checklist refer to item numbers, we kept the existing items under their previous item numbers except for some renumbering of items 2 to 5. We added additional items either as a sub-item under an existing item, an entirely new item number at the end of the checklist, or (with item 3) an interjected item into a renumbered segment. We have summarised the noteworthy general changes in box 1 and specific changes in box 2. The CONSORT website contains a side by side comparison of the 2001 and 2010 versions.

Implications and limitations

We developed CONSORT 2010 to assist authors in writing reports of randomised controlled trials, editors and peer reviewers in reviewing manuscripts for publication, and readers in critically appraising published articles. The CONSORT 2010 Explanation and Elaboration provides elucidation and context to the checklist items. We strongly recommend using the explanation and elaboration in conjunction with the checklist to foster complete, clear, and transparent reporting and aid appraisal of published trial reports.

CONSORT 2010 focuses predominantly on the two group, parallel randomised controlled trial, which accounts for over half of trials in the literature.[2] Most of the items from the CONSORT 2010 Statement, however, pertain to all types of randomised trials. Nevertheless, some types of trials or trial situations dictate the need for additional information in the trial report. When in doubt, authors, editors, and readers should consult the CONSORT website for any CONSORT extensions, expansions (amplifications), implementations, or other guidance that may be relevant.

The evidence based approach we have used for CONSORT also served as a model for development of other reporting guidelines, such as for reporting systematic reviews and meta-analyses of studies evaluating interventions,[16] diagnostic studies,[17] and observational studies.[18] The explicit goal of all these initiatives is to improve reporting. The Enhancing the Quality and Transparency of Health Research (EQUATOR) Network will facilitate development of reporting guidelines and help disseminate the guidelines: www.equator-network.org provides information on all reporting guidelines in health research.

With CONSORT 2010, we again intentionally declined to produce a rigid structure for the reporting of randomised trials. Indeed, SORT[19] tried a rigid format, and it failed in a pilot run with an editor and authors.[20] Consequently, the format of articles should abide by journal style, editorial directions, the traditions of the research field addressed, and, where possible, author preferences. We do not wish to standardise the structure of reporting. Authors should simply address checklist items somewhere in the article, with ample detail and lucidity. That stated, we think that manuscripts benefit from frequent subheadings within the major sections, especially the methods and results sections.

Box 2 | Noteworthy specific changes in CONSORT 2010 Statement

- *Item 1b (title and abstract)*—We added a sub-item on providing a structured summary of trial design, methods, results, and conclusions and referenced the CONSORT for abstracts article[21]
- *Item 2b (introduction)*—We added a new sub-item (formerly item 5 in CONSORT 2001) on "Specific objectives or hypotheses"
- *Item 3a (trial design)*—We added a new item including this sub-item to clarify the basic trial design (such as parallel group, crossover, cluster) and the allocation ratio
- *Item 3b (trial design)*—We added a new sub-item that addresses any important changes to methods after trial commencement, with a discussion of reasons
- *Item 4 (participants)*—Formerly item 3 in CONSORT 2001
- *Item 5 (interventions)*—Formerly item 4 in CONSORT 2001. We encouraged greater specificity by stating that descriptions of interventions should include "sufficient details to allow replication"[3]
- *Item 6 (outcomes)*—We added a sub-item on identifying any changes to the primary and secondary outcome (endpoint) measures after the trial started. This followed from empirical evidence that authors frequently provide analyses of outcomes in their published papers that were not the prespecified primary and secondary outcomes in their protocols, while ignoring their prespecified outcomes (that is, selective outcome reporting).[4 22] We eliminated text on any methods used to enhance the quality of measurements
- *Item 9 (allocation concealment mechanism)*—We reworded this to include mechanism in both the report topic and the descriptor to reinforce that authors should report the actual steps taken to ensure allocation concealment rather than simply report imprecise, perhaps banal, assurances of concealment
- *Item 11 (blinding)*—We added the specification of how blinding was done and, if relevant, a description of the similarity of interventions and procedures. We also eliminated text on "how the success of blinding (masking) was assessed" because of a lack of empirical evidence supporting the practice as well as theoretical concerns about the validity of any such assessment[23 24]
- *Item 12a (statistical methods)*—We added that statistical methods should also be provided for analysis of secondary outcomes
- *Sub-item 14b (recruitment)*—Based on empirical research, we added a sub-item on "Why the trial ended or was stopped"[25]
- *Item 15 (baseline data)*—We specified "A table" to clarify that baseline and clinical characteristics of each group are most clearly expressed in a table
- *Item 16 (numbers analysed)*—We replaced mention of "intention to treat" analysis, a widely misused term, by a more explicit request for information about retaining participants in their original assigned groups[26]
- *Sub-item 17b (outcomes and estimation)*—For appropriate clinical interpretability, prevailing experience suggested the addition of "For binary outcomes, presentation of both relative and absolute effect sizes is recommended"[27]
- *Item 19 (harms)*—We included a reference to the CONSORT paper on harms[28]
- *Item 20 (limitations)*—We changed the topic from "Interpretation" and supplanted the prior text with a sentence focusing on the reporting of sources of potential bias and imprecision
- *Item 22 (interpretation)*—We changed the topic from "Overall evidence." Indeed, we understand that authors should be allowed leeway for interpretation under this nebulous heading. However, the CONSORT Group expressed concerns that conclusions in papers frequently misrepresented the actual analytical results and that harms were ignored or marginalised. Therefore, we changed the checklist item to include the concepts of results matching interpretations and of benefits being balanced with harms
- *Item 23 (registration)*—We added a new item on trial registration. Empirical evidence supports the need for trial registration, and recent requirements by journal editors have fostered compliance[29]
- *Item 24 (protocol)*—We added a new item on availability of the trial protocol. Empirical evidence suggests that authors often ignore, in the conduct and reporting of their trial, what they stated in the protocol.[4 22] Hence, availability of the protocol can instigate adherence to the protocol before publication and facilitate assessment of adherence after publication
- *Item 25 (funding)*—We added a new item on funding. Empirical evidence points toward funding source sometimes being associated with estimated treatment effects[30]

CONSORT urges completeness, clarity, and transparency of reporting, which simply reflects the actual trial design and conduct. However, as a potential drawback, a reporting guideline might encourage some authors to report fictitiously the information suggested by the guidance rather than what was actually done. Authors, peer reviewers, and editors should vigilantly guard against that potential drawback and refer, for example, to trial protocols, to information on trial registers, and to regulatory agency websites. Moreover, the CONSORT 2010 Statement does not include recommendations for designing and conducting randomised trials. The items should elicit clear pronouncements of how and what the authors did, but do not contain any judgments on how and what the authors should have done. Thus, CONSORT 2010 is not intended as an instrument to evaluate the quality of a trial. Nor is it appropriate to use the checklist to construct a "quality score."

Nevertheless, we suggest that researchers begin trials with their end publication in mind. Poor reporting allows authors, intentionally or inadvertently, to escape scrutiny of any weak aspects of their trials. However, with wide adoption of CONSORT by journals and editorial groups, most authors should have to report transparently all important aspects of their trial. The ensuing scrutiny rewards well conducted trials and penalises poorly conducted trials. Thus, investigators should understand the CONSORT 2010 reporting guidelines before starting a trial as a further incentive to design and conduct their trials according to rigorous standards.

CONSORT 2010 supplants the prior version published in 2001. Any support for the earlier version accumulated from journals or editorial groups will automatically extend to this newer version, unless specifically requested otherwise. Journals that do not currently support CONSORT may do so by registering on the CONSORT website. If a journal supports or endorses CONSORT 2010, it should cite one of the original versions of CONSORT 2010, the CONSORT 2010 Explanation and Elaboration, and the CONSORT website in their "Instructions to authors." We suggest that authors who wish to cite CONSORT should cite this or another of the original journal versions of CONSORT 2010 Statement, and, if appropriate, the CONSORT 2010 Explanation and Elaboration.[13] All CONSORT material can be accessed through the original publishing journals or the CONSORT website. Groups or individuals who desire to translate the CONSORT 2010 Statement into

bmj.com: recent Research Methods & Reporting articles

- Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design (2010;340:c1066)
- Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes (2010;340:b5087)
- Economic impact of disease and injury: counting what matters (2010;340:c924)
- The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews (2010;340:c365)
- Meta-analysis of individual participant data: rationale, conduct, and reporting (2010;340:c221)

other languages should first consult the CONSORT policy statement on the website.

We emphasise that CONSORT 2010 represents an evolving guideline. It requires perpetual reappraisal and, if necessary, modifications. In the future we will further revise the CONSORT material considering comments, criticisms, experiences, and accumulating new evidence. We invite readers to submit recommendations via the CONSORT website.

Author contributions: KFS, DM, and DGA participated in meetings and regular conference calls, planned the CONSORT 2007 meeting at Montebello, developed the agenda, prepared background research, identified and invited participants, contributed to the CONSORT meeting, drafted the manuscript, and, after critical review by the CONSORT Group, finalised the text of the manuscript. Members of the CONSORT Group attended the meeting, except for those noted below, and provided input on and review of the revised checklist and text of this article. Some members also prepared background material.

The CONSORT Group contributors to CONSORT 2010: DG Altman, Centre for Statistics in Medicine, University of Oxford; Virginia Barbour, *PLoS Medicine*; Jesse A Berlin, Johnson & Johnson Pharmaceutical Research and Development, USA; Isabelle Boutron, University Paris 7 Denis Diderot, Assistance Publique des Hôpitaux de Paris, INSERM, France; PJ Devereaux, McMaster University, Canada; Kay Dickersin, Johns Hopkins Bloomberg School of Public Health, USA; Diana Elbourne, London School of Hygiene & Tropical Medicine; Susan Ellenberg, University of Pennsylvania School of Medicine, USA; Val Gebski, University of Sydney, Australia; Steven Goodman, *Journal of the Society for Clinical Trials*, USA; Peter C Gøtzsche, Nordic Cochrane Centre, Denmark; Trish Groves, *BMJ*; Steven Grunberg, American Society of Clinical Oncology, USA; Brian Haynes, McMaster University, Canada; Sally Hopewell, Centre for Statistics in Medicine, University of Oxford; Astrid James, *Lancet*; Peter Juhn, Johnson & Johnson, USA; Philippa Middleton, University of Adelaide, Australia; Don Minckler, University of California Irvine, USA; D Moher, Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Canada; Victor M Montori, Knowledge and Encounter Research Unit, Mayo Clinic College of Medicine, USA; Cynthia Mulrow, *Annals of Internal Medicine*, USA; Stuart Pocock, London School of Hygiene & Tropical Medicine; Drummond Rennie, *JAMA*, USA; David L Schriger, *Annals of Emergency Medicine*, USA; KF Schulz, Family Health International, USA; Iveta Simera, EQUATOR Network; Elizabeth Wager, Sideview.

Contributors to CONSORT 2010 who did not attend the Montebello meeting: Mike Clarke, UK Cochrane Centre; Gordon Guyatt, McMaster University, Canada.

Funding: We received financial support from United Kingdom National Institute for Health Research and the Medical Research Council; Canadian Institutes of Health Research; Presidents Fund, Canadian Institutes of Health Research; Johnson & Johnson; *BMJ*; and the American Society for Clinical Oncology. DGA is supported by Cancer Research UK, DM by a University of Ottawa Research Chair, and KFS by Family Health International. None of the sponsors had any involvement in the planning, execution, or writing of the CONSORT documents. Additionally, no funder played a role in drafting the manuscript.

Competing interests: Uniform disclosure of potential conflicts of interest: all authors have completed the ICMJE unified competing interest form at www.icmje.org/coi_disclosure.pdf (available from the corresponding author) and declare (1) DM received grants for this work from Johnson & Johnson, *BMJ*, and American Society for Clinical Oncology; KFS and DGA received support for travel to meetings for this work from Johnson & Johnson, *BMJ*, and American Society for Clinical Oncology; (2) KFS and DA had travel expenses reimbursed by the EQUATOR Network; KFS has received honoraria for delivering educational presentations for the American Board of Obstetrics and Gynecology Foundation for Excellence in Women's Health Care, Ortho-McNeil Janssen Scientific Affairs, and the American College of Obstetrics and Gynecology; and has done consultancy for Wyeth. All authors also declare (3) no spouses, partners, or children with relationships with commercial entities that might have an interest in the submitted work; (4) no non-financial interests that may be relevant to the submitted work.

1 Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
2 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.
3 Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;336:1472-4.
4 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.
5 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
6 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
7 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134:657-62.
8 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.
9 Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263-7.
10 Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: a comparative study of articles indexed by PubMed. *BMJ* 2010;340:c723.
11 Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328:702-8.
12 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152-60.
13 Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
14 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
15 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
16 Moher D, Liberati A, Tetzlaff J, Altman DG, for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
17 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41-4.
18 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, for the STROBE Initiative. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806-8.
19 Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994;272:1926-31.
20 Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers. *JAMA* 1995;273:1054-5.
21 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008;371:281-3.
22 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
23 Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol* 2007;36:664-5.
24 Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;359:696-700.
25 Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.
26 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.
27 Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA* 2002;287:2813-4.
28 Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
29 De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Lancet* 2004;364:911-2.
30 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
31 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.
32 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295-309.
33 Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 2006;144:364-7.
34 Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.

RESEARCH METHODS & REPORTING

CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials

David Moher,[1] Sally Hopewell,[2] Kenneth F Schulz,[3] Victor Montori,[4] Peter C Gøtzsche,[5] P J Devereaux,[6] Diana Elbourne,[7] Matthias Egger,[8] Douglas G Altman[2]

[1]Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa Hospital, Ottawa, Ontario, Canada, K1H 8L6
[2]Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford
[3]Family Health International, Research Triangle Park, NC 27709, USA
[4]UK Knowledge and Encounter Research Unit, Mayo Clinic, Rochester, MN, USA
[5]The Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, Copenhagen, Denmark
[6]McMaster University Health Sciences Centre, Hamilton, Canada
[7]Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London
[8]Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

Correspondence to: D Moher
dmoher@ohri.ca

Accepted: 8 February 2010

Cite this as: *BMJ* 2010;340:c869
doi: 10.1136/bmj.c869

ABSTRACT

Overwhelming evidence shows the quality of reporting of randomised controlled trials (RCTs) is not optimal. Without transparent reporting, readers cannot judge the reliability and validity of trial findings nor extract information for systematic reviews. Recent methodological analyses indicate that inadequate reporting and design are associated with biased estimates of treatment effects. Such systematic error is seriously damaging to RCTs, which are considered the gold standard for evaluating interventions because of their ability to minimise or avoid bias.

A group of scientists and editors developed the CONSORT (Consolidated Standards of Reporting Trials) statement to improve the quality of reporting of RCTs. It was first published in 1996 and updated in 2001. The statement consists of a checklist and flow diagram that authors can use for reporting an RCT. Many leading medical journals and major international editorial groups have endorsed the CONSORT statement. The statement facilitates critical appraisal and interpretation of RCTs.

During the 2001 CONSORT revision, it became clear that explanation and elaboration of the principles underlying the CONSORT statement would help investigators and others to write or appraise trial reports. A CONSORT explanation and elaboration article was published in 2001 alongside the 2001 version of the CONSORT statement.

After an expert meeting in January 2007, the CONSORT statement has been further revised and is published as the CONSORT 2010 Statement. This update improves the wording and clarity of the previous checklist and incorporates recommendations related to topics that have only recently received recognition, such as selective outcome reporting bias.

This explanatory and elaboration document—intended to enhance the use, understanding, and dissemination of the CONSORT statement—has also been extensively revised. It presents the meaning and rationale for each new and updated checklist item providing examples of good reporting and, where possible, references to relevant empirical studies. Several examples of flow diagrams are included.

The CONSORT 2010 Statement, this revised explanatory and elaboration document, and the associated website (www.consort-statement.org) should be helpful resources to improve reporting of randomised trials.

"The whole of medicine depends on the transparent reporting of clinical trials."[1]

Well designed and properly executed randomised controlled trials (RCTs) provide the most reliable evidence on the efficacy of healthcare interventions, but trials with inadequate methods are associated with bias, especially exaggerated treatment effects.[2-5] Biased results from poorly designed and reported trials can mislead decision making in health care at all levels, from treatment decisions for a patient to formulation of national public health policies.

Critical appraisal of the quality of clinical trials is possible only if the design, conduct, and analysis of RCTs are thoroughly and accurately described in the report. Far from being transparent, the reporting of RCTs is often incomplete,[6-9] compounding problems arising from poor methodology.[10-15]

Incomplete and inaccurate reporting

Many reviews have documented deficiencies in reports of clinical trials. For example, information on the method used in a trial to assign participants to comparison groups was reported in only 21% of 519 trial reports indexed in PubMed in 2000,[16] and only 34% of 616 reports indexed in 2006.[17] Similarly, only 45% of trial reports indexed in PubMed in 2000[16] and 53% in 2006[17] defined a primary end point, and only 27% in 2000 and 45% in 2006 reported a sample size calculation. Reporting is not only often incomplete but also sometimes inaccurate. Of 119 reports stating that all participants were included in the analysis in the groups to which they were originally assigned (intention-to-treat analysis), 15 (13%) excluded patients or did not analyse all patients as allocated.[18] Many other reviews have found that inadequate reporting is common in specialty journals[16 19] and journals published in languages other than English.[20 21]

Proper randomisation reduces selection bias at trial entry and is the crucial component of high quality RCTs.[22] Successful randomisation hinges on two steps: generation of an unpredictable allocation sequence and concealment of this sequence from the investigators enrolling participants (see box 1).[2 23]

Unfortunately, despite that central role, reporting of the methods used for allocation of participants to interventions is also generally inadequate. For example, 5% of 206 reports of supposed RCTs in obstetrics and gynaecology journals described studies that were not truly randomised.[23] This estimate is conservative, as most reports do not at present provide adequate information about the method of allocation.[20 23 30-33]

Box 1 | Treatment allocation. What's so special about randomisation?

The method used to assign interventions to trial participants is a crucial aspect of clinical trial design. Random assignment is the preferred method; it has been successfully used regularly in trials for more than 50 years.[24] Randomisation has three major advantages.[25] First, when properly implemented, it eliminates selection bias, balancing both known and unknown prognostic factors, in the assignment of treatments. Without randomisation, treatment comparisons may be prejudiced, whether consciously or not, by selection of participants of a particular kind to receive a particular treatment. Second, random assignment permits the use of probability theory to express the likelihood that any difference in outcome between intervention groups merely reflects chance.[26] Third, random allocation, in some situations, facilitates blinding the identity of treatments to the investigators, participants, and evaluators, possibly by use of a placebo, which reduces bias after assignment of treatments.[27] Of these three advantages, reducing selection bias at trial entry is usually the most important.[28]

Successful randomisation in practice depends on two interrelated aspects—adequate generation of an unpredictable allocation sequence and concealment of that sequence until assignment occurs.[2 23] A key issue is whether the schedule is known or predictable by the people involved in allocating participants to the comparison groups.[29] The treatment allocation system should thus be set up so that the person enrolling participants does not know in advance which treatment the next person will get, a process termed allocation concealment.[2 23] Proper allocation concealment shields knowledge of forthcoming assignments, whereas proper random sequences prevent correct anticipation of future assignments based on knowledge of past assignments.

Improving the reporting of RCTs: the CONSORT statement

DerSimonian and colleagues suggested that "editors could greatly improve the reporting of clinical trials by providing authors with a list of items that they expected to be strictly reported."[34] Early in the 1990s, two groups of journal editors, trialists, and methodologists independently published recommendations on the reporting of trials.[35 36] In a subsequent editorial, Rennie urged the two groups to meet and develop a common set of recommendations [37]; the outcome was the CONSORT statement (Consolidated Standards of Reporting Trials).[38]

The CONSORT statement (or simply CONSORT) comprises a checklist of essential items that should be included in reports of RCTs and a diagram for documenting the flow of participants through a trial. It is aimed at primary reports of RCTs with two group, parallel designs. Most of CONSORT is also relevant to a wider class of trial designs, such as non-inferiority, equivalence, factorial, cluster, and crossover trials. Extensions to the CONSORT checklist for reporting trials with some of these designs have been published,[39-41] as have those for reporting certain types of data (harms [42]), types of interventions (non-pharmacological treatments [43], herbal interventions[44]), and abstracts.[45]

The objective of CONSORT is to provide guidance to authors about how to improve the reporting of their trials. Trial reports need be clear, complete, and transparent. Readers, peer reviewers, and editors can also use CONSORT to help them critically appraise and interpret reports of RCTs. However, CONSORT was not meant to be used as a quality assessment instrument. Rather, the content of CONSORT focuses on items related to the internal and external validity of trials. Many items not explicitly mentioned in CONSORT should also be included in a report, such as information about approval by an ethics committee, obtaining informed consent from participants, and, where relevant, existence of a data safety and monitoring committee. In addition, any other aspects of a trial that are mentioned should be properly reported, such as information pertinent to cost effectiveness analysis.[46-48]

Since its publication in 1996, CONSORT has been supported by more than 400 journals (www.consort-statement.org) and several editorial groups, such as the International Committee of Medical Journal Editors.[49] The introduction of CONSORT within journals is associated with improved quality of reports of RCTs.[17 50 51] However, CONSORT is an ongoing initiative, and the CONSORT statement is revised periodically.[3] CONSORT was last revised nine years ago, in 2001.[52-54] Since then the evidence base to inform CONSORT has grown considerably; empirical data have highlighted new concerns regarding the reporting of RCTs, such as selective outcome reporting.[55-57] A CONSORT Group meeting was therefore convened in January 2007, in Canada, to revise the 2001 CONSORT statement and its accompanying explanation and elaboration document. The revised checklist is shown in table 1 and the flow diagram, not revised, in fig 1.[52-54]

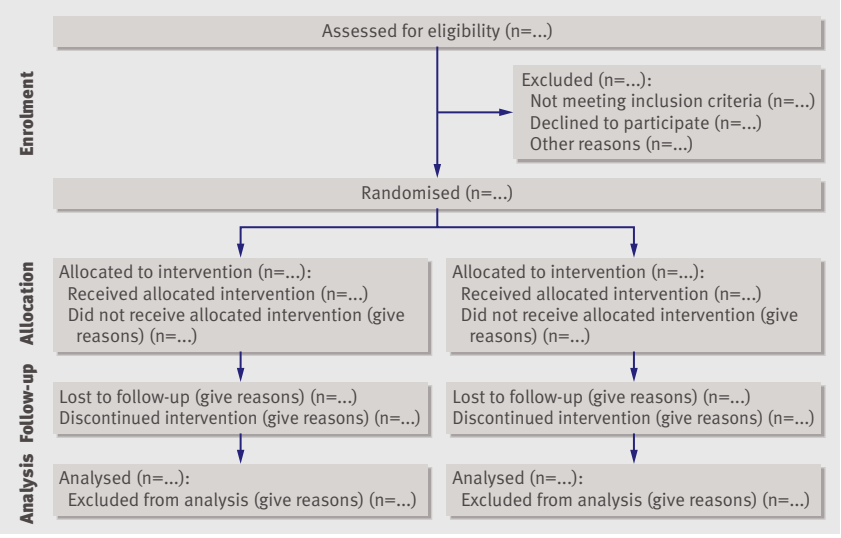


Fig 1 | Flow diagram of the progress through the phases of a parallel randomised trial of two groups (that is, enrolment, intervention allocation, follow-up, and data analysis)[52-54]

The CONSORT 2010 Statement: explanation and elaboration

During the 2001 CONSORT revision, it became clear that explanation and elaboration of the principles underlying the CONSORT statement would help investigators and others to write or appraise trial reports. The CONSORT explanation and elaboration article[58] was published in 2001 alongside the 2001 version of the CONSORT statement. It discussed the rationale and scientific background for each item and provided published examples of good reporting. The rationale for revising that article is similar to that for revising the statement, described above. We briefly describe below the main additions and deletions to this version of the explanation and elaboration article.

The CONSORT 2010 Explanation and Elaboration: changes

We have made several substantive and some cosmetic changes to this version of the CONSORT explanatory document (full details are highlighted in the 2010 version of the CONSORT statement[59]). Some reflect changes to the CON-

Table 1 | **CONSORT 2010 checklist of information to include when reporting a randomised trial***

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts[45 65])	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[42])	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials,[40] non-inferiority and equivalence trials,[39] non-pharmacological treatments,[43] herbal interventions,[44] and pragmatic trials.[41] Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

SORT checklist; there are three new checklist items in the CONSORT 2010 checklist—such as item 24, which asks authors to report where their trial protocol can be accessed. We have also updated some existing explanations, including adding more recent references to methodological evidence, and used some better examples. We have removed the glossary, which is now available on the CONSORT website (www.consort-statement.org). Where possible, we describe the findings of relevant empirical studies. Many excellent books on clinical trials offer fuller discussion of methodological issues.[60-62] Finally, for convenience, we sometimes refer to "treatments" and "patients," although we recognise that not all interventions evaluated in RCTs are treatments and not all participants are patients.

Table 2 | Items to include when reporting a randomised trial in a journal abstract

Item	Description
Authors	Contact details for the corresponding author
Trial design	Description of the trial design (such as parallel, cluster, non-inferiority)
Methods:	
Participants	Eligibility criteria for participants and the settings where the data were collected
Interventions	Interventions intended for each group
Objective	Specific objective or hypothesis
Outcome	Clearly defined primary outcome for this report
Randomisation	How participants were allocated to interventions
Blinding (masking)	Whether participants, care givers, and those assessing the outcomes were blinded to group assignment
Results:	
Numbers randomised	Number of participants randomised to each group
Recruitment	Trial status
Numbers analysed	Number of participants analysed in each group
Outcome	For the primary outcome, a result for each group and the estimated effect size and its precision
Harms	Important adverse events or side effects
Conclusions	General interpretation of the results
Trial registration	Registration number and name of trial register
Funding	Source of funding

Checklist items

Title and abstract

Item 1a. Identification as a randomised trial in the title.

Example—"Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety."[63]

Explanation—The ability to identify a report of a randomised trial in an electronic database depends to a large extent on how it was indexed. Indexers may not classify a report as a randomised trial if the authors do not explicitly report this information.[64] To help ensure that a study is appropriately indexed and easily identified, authors should use the word "randomised" in the title to indicate that the participants were randomly assigned to their comparison groups.

Item 1b. Structured summary of trial design, methods, results, and conclusions

For specific guidance see CONSORT for abstracts.[45 65]

Explanation—Clear, transparent, and sufficiently detailed abstracts are important because readers often base their assessment of a trial on such information. Some readers use an abstract as a screening tool to decide whether to read the full article. However, as not all trials are freely available and some health professionals do not have access to the full trial reports, healthcare decisions are sometimes made on the basis of abstracts of randomised trials.[66]

A journal abstract should contain sufficient information about a trial to serve as an accurate record of its conduct and findings, providing optimal information about the trial within the space constraints and format of a journal. A properly constructed and written abstract helps individuals to assess quickly the relevance of the findings and aids the retrieval of relevant reports from electronic databases.[67] The abstract should accurately reflect what is included in the full journal article and should not include information that does not appear in the body of the paper. Studies comparing the accuracy of information reported in a journal abstract with that reported in the text of the full publication have found claims that are inconsistent with, or missing from, the body of the full article.[68-71] Conversely, omitting important harms from the abstract could seriously mislead someone's interpretation of the trial findings.[42 72]

A recent extension to the CONSORT statement provides a list of essential items that authors should include when reporting the main results of a randomised trial in a journal (or conference) abstract (see table 2).[45] We strongly recommend the use of structured abstracts for reporting randomised trials. They provide readers with information about the trial under a series of headings pertaining to the design, conduct, analysis, and interpretation.[73] Some studies have found that structured abstracts are of higher quality than the more traditional descriptive abstracts[74 75] and that they allow readers to find information more easily.[76] We recognise that many journals have developed their own structure and word limit for reporting abstracts. It is not our intention to suggest changes to these formats, but to recommend what information should be reported.

Introduction

Item 2a. Scientific background and explanation of rationale

Example—"Surgery is the treatment of choice for patients with disease stage I and II non-small cell lung cancer (NSCLC) ... An NSCLC meta-analysis combined the results from eight randomised trials of surgery versus surgery plus adjuvant cisplatin-based chemotherapy and showed a small, but not significant (p=0.08), absolute survival benefit of around 5% at 5 years (from 50% to 55%). At the time the current trial was designed (mid-1990s), adjuvant chemotherapy had not become standard clinical practice ... The clinical rationale for neo-adjuvant chemotherapy is three-fold: regression of the primary cancer could be achieved thereby facilitating and simplifying or reducing subsequent surgery; undetected micro-metastases could be dealt with at the start of treatment; and there might be inhibition of the putative stimulus to residual cancer by growth factors released by surgery and by subsequent wound healing ... The current trial was therefore set up to compare, in patients with resectable NSCLC, surgery alone versus three cycles of platinum-based chemotherapy followed by surgery in terms of overall survival, quality of life, pathological staging, resectability rates, extent of surgery, and time to and site of relapse."[77]

Explanation—Typically, the introduction consists of free flowing text, in which authors explain the scientific background and rationale for their trial, and its general outline. It may also be appropriate to include here the objectives of the trial (see item 2b).The rationale may be explanatory (for example, to assess the possible influence of a drug on renal function) or pragmatic (for example, to guide practice by comparing the benefits and harms of two treatments). Authors should report any evidence of the benefits and harms of active interventions included in a trial and should suggest a plausible explanation for how the interventions might work, if this is not obvious.[78]

The Declaration of Helsinki states that biomedical research involving people should be based on a thorough knowledge of the scientific literature.[79] That is, it is unethical to expose humans unnecessarily to the risks of research. Some clinical trials have been shown to have been unnecessary because the question they addressed had been or could have been answered by a systematic review of the existing

literature.[80 81] Thus, the need for a new trial should be justified in the introduction. Ideally, it should include a reference to a systematic review of previous similar trials or a note of the absence of such trials.[82]

Item 2b. Specific objectives or hypotheses

Example—"In the current study we tested the hypothesis that a policy of active management of nulliparous labour would: 1. reduce the rate of caesarean section, 2. reduce the rate of prolonged labour; 3. not influence maternal satisfaction with the birth experience."[83]

Explanation—Objectives are the questions that the trial was designed to answer. They often relate to the efficacy of a particular therapeutic or preventive intervention. Hypotheses are pre-specified questions being tested to help meet the objectives. Hypotheses are more specific than objectives and are amenable to explicit statistical evaluation. In practice, objectives and hypotheses are not always easily differentiated. Most reports of RCTs provide adequate information about trial objectives and hypotheses.[84]

Methods

Item 3a. Description of trial design (such as parallel, factorial) including allocation ratio

Example—"This was a multicenter, stratified (6 to 11 years and 12 to 17 years of age, with imbalanced randomisation [2:1]), double-blind, placebo-controlled, parallel-group study conducted in the United States (41 sites)."[85]

Explanation—The word "design" is often used to refer to all aspects of how a trial is set up, but it also has a narrower interpretation. Many specific aspects of the broader trial design, including details of randomisation and blinding, are addressed elsewhere in the CONSORT checklist. Here we seek information on the type of trial, such as parallel group or factorial, and the conceptual framework, such as superiority or non-inferiority, and other related issues not addressed elsewhere in the checklist.

The CONSORT statement focuses mainly on trials with participants individually randomised to one of two "parallel" groups. In fact, little more than half of published trials have such a design.[16] The main alternative designs are multi-arm parallel, crossover, cluster,[40] and factorial designs. Also, most trials are set to identify the superiority of a new intervention, if it exists, but others are designed to assess non-inferiority or equivalence.[39] It is important that researchers clearly describe these aspects of their trial, including the unit of randomisation (such as patient, GP practice, lesion). It is desirable also to include these details in the abstract (see item 1b).

If a less common design is employed, authors are encouraged to explain their choice, especially as such designs may imply the need for a larger sample size or more complex analysis and interpretation.

Although most trials use equal randomisation (such as 1:1 for two groups), it is helpful to provide the allocation ratio explicitly. For drug trials, specifying the phase of the trial (I-IV) may also be relevant.

Item 3b. Important changes to methods after trial commencement (such as eligibility criteria), with reasons

Example—"Patients were randomly assigned to one of six parallel groups, initially in 1:1:1:1:1:1 ratio, to receive either one of five otamixaban ... regimens ... or an active control of unfractionated heparin ... an independent Data Monitoring Committee reviewed unblinded data for patient safety; no interim analyses for efficacy or futility were done. During the trial, this committee recommended that the group receiving the lowest dose of otamixaban (0·035 mg/kg/h) be discontinued because of clinical evidence of inadequate anticoagulation. The protocol was immediately amended in accordance with that recommendation, and participants were subsequently randomly assigned in 2:2:2:2:1 ratio to the remaining otamixaban and control groups, respectively."[86]

Explanation—A few trials may start without any fixed plan (that is, are entirely exploratory), but the most will have a protocol that specifies in great detail how the trial will be conducted. There may be deviations from the original protocol, as it is impossible to predict every possible change in circumstances during the course of a trial. Some trials will therefore have important changes to the methods after trial commencement.

Changes could be due to external information becoming available from other studies, or internal financial difficulties, or could be due to a disappointing recruitment rate. Such protocol changes should be made without breaking the blinding on the accumulating data on participants' outcomes. In some trials, an independent data monitoring committee will have as part of its remit the possibility of recommending protocol changes based on seeing unblinded data. Such changes might affect the study methods (such as changes to treatment regimens, eligibility criteria, randomisation ratio, or duration of follow-up) or trial conduct (such as dropping a centre with poor data quality).[87]

Some trials are set up with a formal "adaptive" design. There is no universally accepted definition of these designs, but a working definition might be "a multistage study design that uses accumulating data to decide how to modify aspects of the study without undermining the validity and integrity of the trial."[88] The modifications are usually to the sample sizes and the number of treatment arms and can lead to decisions being made more quickly and with more efficient use of resources. There are, however, important ethical, statistical, and practical issues in considering such a design.[89 90]

Whether the modifications are explicitly part of the trial design or in response to changing circumstances, it is essential that they are fully reported to help the reader interpret the results. Changes from protocols are not currently well reported. A review of comparisons with protocols showed that about half of journal articles describing RCTs had an unexplained discrepancy in the primary outcomes.[57] Frequent unexplained discrepancies have also been observed for details of randomisation, blinding,[91] and statistical analyses.[92]

Item 4a. Eligibility criteria for participants

Example—"Eligible participants were all adults aged 18 or over with HIV who met the eligibility criteria for antiretroviral therapy according to the Malawian national HIV treatment guidelines (WHO clinical stage III or IV or any WHO stage with a CD4 count <250/mm^3) and who were starting treatment with a BMI <18.5. Exclusion criteria were pregnancy

and lactation or participation in another supplementary feeding programme."[93]

Explanation—A comprehensive description of the eligibility criteria used to select the trial participants is needed to help readers interpret the study. In particular, a clear understanding of these criteria is one of several elements required to judge to whom the results of a trial apply—that is, the trial's generalisability (applicability) and relevance to clinical or public health practice (see item 21).[94] A description of the method of recruitment, such as by referral or self selection (for example, through advertisements), is also important in this context. Because they are applied before randomisation, eligibility criteria do not affect the internal validity of a trial, but they are central to its external validity.

Typical and widely accepted selection criteria relate to the nature and stage of the disease being studied, the exclusion of persons thought to be particularly vulnerable to harm from the study intervention, and to issues required to ensure that the study satisfies legal and ethical norms. Informed consent by study participants, for example, is typically required in intervention studies. The common distinction between inclusion and exclusion criteria is unnecessary; the same criterion can be phrased to include or exclude participants.[95]

Despite their importance, eligibility criteria are often not reported adequately. For example, eight published trials leading to clinical alerts by the National Institutes of Health specified an average of 31 eligibility criteria in their protocols, but only 63% of the criteria were mentioned in the journal articles, and only 19% were mentioned in the clinical alerts.[96] Similar deficiencies were found for HIV clinical trials.[97] Among 364 reports of RCTs in surgery, 25% did not specify any eligibility criteria.[98]

Item 4b. Settings and locations where the data were collected

Example—"The study took place at the antiretroviral therapy clinic of Queen Elizabeth Central Hospital in Blantyre, Malawi, from January 2006 to April 2007. Blantyre is the major commercial city of Malawi, with a population of 1 000 000 and an estimated HIV prevalence of 27% in adults in 2004."[93]

Explanation—Along with the eligibility criteria for participants (see item 4a) and the description of the interventions (see item 5), information on the settings and locations is crucial to judge the applicability and generalisability of a trial. Were participants recruited from primary, secondary, or tertiary health care or from the community? Healthcare institutions vary greatly in their organisation, experience, and resources and the baseline risk for the condition under investigation. Other aspects of the setting (including the social, economic, and cultural environment and the climate) may also affect a study's external validity.

Authors should report the number and type of settings and describe the care providers involved. They should report the locations in which the study was carried out, including the country, city if applicable, and immediate environment (for example, community, office practice, hospital clinic, or inpatient unit). In particular, it should be clear whether the trial was carried out in one or several centres ("multicentre trials"). This description should provide enough information so that readers can judge whether the results of the trial could be relevant to their own setting. The environment in which the trial is conducted may differ considerably from the setting in which the trial's results are later used to guide practice and policy.[94 99] Authors should also report any other information about the settings and locations that could have influenced the observed results, such as problems with transportation that might have affected patient participation or delays in administering interventions.

Item 5. The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

Examples—"In POISE, patients received the first dose of the study drug (ie, oral extended-release metoprolol 100 mg or matching placebo) 2-4 h before surgery. Study drug administration required a heart rate of 50 bpm or more and a systolic blood pressure of 100 mm Hg or greater; these haemodynamics were checked before each administration. If, at any time during the first 6 h after surgery, heart rate was 80 bpm or more and systolic blood pressure was 100 mm Hg or higher, patients received their first postoperative dose (extended-release metoprolol 100 mg or matched placebo) orally. If the study drug was not given during the first 6 h, patients received their first postoperative dose at 6 h after surgery. 12 h after the first postoperative dose, patients started taking oral extended-release metoprolol 200 mg or placebo every day for 30 days. If a patient's heart rate was consistently below 45 bpm or their systolic blood pressure dropped below 100 mm Hg, study drug was withheld until their heart rate or systolic blood pressure recovered; the study drug was then restarted at 100 mg once daily. Patients whose heart rate was consistently 45-49 bpm and systolic blood pressure exceeded 100 mm Hg delayed taking the study drug for 12 h."[100]

"Patients were randomly assigned to receive a custom-made neoprene splint to be worn at night or to usual care. The splint was a rigid rest orthosis recommended for use only at night. It covered the base of the thumb and the thenar eminence but not the wrist (Figure 1). Splints were made by 3 trained occupational therapists, who adjusted the splint for each patient so that the first web could be opened and the thumb placed in opposition with the first long finger. Patients were encouraged to contact the occupational therapist if they felt that the splint needed adjustment, pain increased while wearing the splint, or they had adverse effects (such as skin erosion). Because no treatment can be considered the gold standard in this situation, patients in the control and intervention groups received usual care at the discretion of their physician (general practitioner or rheumatologist). We decided not to use a placebo because, to our knowledge, no placebo for splinting has achieved successful blinding of patients, as recommended."[101]

Explanation—Authors should describe each intervention thoroughly, including control interventions. The description should allow a clinician wanting to use the intervention to know exactly how to administer the intervention that was evaluated in the trial.[102] For a drug intervention, information would include the drug name, dose, method of administration (such as oral, intravenous), timing and duration of administration, conditions under which interventions are withheld, and titration regimen if applicable. If the control group is to receive "usual care" it is important to describe thoroughly what that constitutes. If the control group or intervention group is to receive a combination of interven-

tions the authors should provide a thorough description of each intervention, an explanation of the order in which the combination of interventions are introduced or withdrawn, and the triggers for their introduction if applicable.

Specific extensions of the CONSORT statement address the reporting of non-pharmacologic and herbal interventions and their particular reporting requirements (such as expertise, details of how the interventions were standardised).[43 44] We recommend readers consult the statements for non-pharmacologic and herbal interventions as appropriate.

Item 6a. Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed

Example—"The primary endpoint with respect to efficacy in psoriasis was the proportion of patients achieving a 75% improvement in psoriasis activity from baseline to 12 weeks as measured by the PASI [psoriasis area and severity index] Additional analyses were done on the percentage change in PASI scores and improvement in target psoriasis lesions."[103]

Explanation—All RCTs assess response variables, or outcomes (end points), for which the groups are compared. Most trials have several outcomes, some of which are of more interest than others. The primary outcome measure is the pre-specified outcome considered to be of greatest importance to relevant stakeholders (such a patients, policy makers, clinicians, funders) and is usually the one used in the sample size calculation (see item 7). Some trials may have more than one primary outcome. Having several primary outcomes, however, incurs the problems of interpretation associated with multiplicity of analyses (see items 18 and 20) and is not recommended. Primary outcomes should be explicitly indicated as such in the report of an RCT. Other outcomes of interest are secondary outcomes (additional outcomes). There may be several secondary outcomes, which often include unanticipated or unintended effects of the intervention (see item 19), although harms should always be viewed as important whether they are labelled primary or secondary.

All outcome measures, whether primary or secondary, should be identified and completely defined. The principle here is that the information provided should be sufficient to allow others to use the same outcomes.[102] When outcomes are assessed at several time points after randomisation, authors should also indicate the pre-specified time point of primary interest. For many non-pharmacological interventions it is helpful to specify who assessed outcomes (for example, if special skills are required to do so) and how many assessors there were.[43]

Where available and appropriate, the use of previously developed and validated scales or consensus guidelines should be reported,[104 105] both to enhance quality of measurement and to assist in comparison with similar studies.[106] For example, assessment of quality of life is likely to be improved by using a validated instrument.[107] Authors should indicate the provenance and properties of scales.

More than 70 outcomes were used in 196 RCTs of non-steroidal anti-inflammatory drugs for rheumatoid arthritis,[108] and 640 different instruments had been used in 2000 trials in schizophrenia, of which 369 had been used only once.[33] Investigation of 149 of those 2000 trials showed that unpublished scales were a source of bias. In non-pharmacological trials, a third of the claims of treatment superiority based on unpublished scales would not have been made if a published scale had been used.[109] Similar data have been reported elsewhere.[110 111] Only 45% of a cohort of 519 RCTs published in 2000 specified the primary outcome[16]; this compares with 53% for a similar cohort of 614 RCTs published in 2006.[17]

Item 6b. Any changes to trial outcomes after the trial commenced, with reasons

Example—"The original primary endpoint was all-cause mortality, but, during a masked analysis, the data and safety monitoring board noted that overall mortality was lower than had been predicted and that the study could not be completed with the sample size and power originally planned. The steering committee therefore decided to adopt co-primary endpoints of all-cause mortality (the original primary endpoint), together with all-cause mortality or cardiovascular hospital admissions (the first pre-specified secondary endpoint)."[112]

Explanation—There are many reasons for departures from the initial study protocol (see item 24). Authors should report all major changes to the protocol, including unplanned changes to eligibility criteria, interventions, examinations, data collection, methods of analysis, and outcomes. Such information is not always reported.

As indicated earlier (see item 6a), most trials record multiple outcomes, with the risk that results will be reported for only a selected subset (see item 17). Pre-specification and reporting of primary and secondary outcomes (see item 6a) should remove such a risk. In some trials, however, circumstances require a change in the way an outcome is assessed or even, as in the example above, a switch to a different outcome. For example, there may be external evidence from other trials or systematic reviews suggesting the end point might not be appropriate, or recruitment or the overall event rate in the trial may be lower than expected.[112] Changing an end point based on unblinded data is much more problematic, although it may be specified in the context of an adaptive trial design.[88] Authors should identify and explain any such changes. Likewise, any changes after the trial began of the designation of outcomes as primary or secondary should be reported and explained.

A comparison of protocols and publications of 102 randomised trials found that 62% of trials reports had at least one primary outcome that was changed, introduced, or omitted compared with the protocol.[55] Primary outcomes also differed between protocols and publications for 40% of a cohort of 48 trials funded by the Canadian Institutes of Health Research.[113] Not one of the subsequent 150 trial reports mentioned, let alone explained, changes from the protocol. Similar results from other studies have been reported recently in a systematic review of empirical studies examining outcome reporting bias.[57]

Item 7a. How sample size was determined

Examples—"To detect a reduction in PHS (postoperative hospital stay) of 3 days (SD 5 days), which is in agreement with the study of Lobo et al[17] with a two-sided 5% significance level and a power of 80%, a sample size of 50 patients per group was necessary, given an anticipated

dropout rate of 10%. To recruit this number of patients a 12-month inclusion period was anticipated."[114]

"Based on an expected incidence of the primary composite endpoint of 11% at 2.25 years in the placebo group, we calculated that we would need 950 primary endpoint events and a sample size of 9650 patients to give 90% power to detect a significant difference between ivabradine and placebo, corresponding to a 19% reduction of relative risk (with a two-sided type 1 error of 5%). We initially designed an event-driven trial, and planned to stop when 950 primary endpoint events had occurred. However, the incidence of the primary endpoint was higher than predicted, perhaps because of baseline characteristics of the recruited patients, who had higher risk than expected (e.g., lower proportion of NYHA class I and higher rates of diabetes and hypertension). We calculated that when 950 primary endpoint events had occurred, the most recently included patients would only have been treated for about 3 months. Therefore, in January 2007, the executive committee decided to change the study from being event-driven to time-driven, and to continue the study until the patients who were randomised last had been followed up for 12 months. This change did not alter the planned study duration of 3 years."[115]

Explanation—For scientific and ethical reasons, the sample size for a trial needs to be planned carefully, with a balance between medical and statistical considerations. Ideally, a study should be large enough to have a high probability (power) of detecting as statistically significant a clinically important difference of a given size if such a difference exists. The size of effect deemed important is inversely related to the sample size necessary to detect it; that is, large samples are necessary to detect small differences. Elements of the sample size calculation are (1) the estimated outcomes in each group (which implies the clinically important target difference between the intervention groups); (2) the α (type I) error level; (3) the statistical power (or the β (type II) error level); and (4), for continuous outcomes, the standard deviation of the measurements.[116] The interplay of these elements and their reporting will differ for cluster trials[40] and non-inferiority and equivalence trials.[39]

Authors should indicate how the sample size was determined. If a formal power calculation was used, the authors should identify the primary outcome on which the calculation was based (see item 6a), all the quantities used in the calculation, and the resulting target sample size per study group. It is preferable to quote the expected result in the control group and the difference between the groups one would not like to overlook. Alternatively, authors could present the percentage with the event or mean for each group used in their calculations. Details should be given of any allowance made for attrition or non-compliance during the study.

Some methodologists have written that so called underpowered trials may be acceptable because they could ultimately be combined in a systematic review and meta-analysis,[117-119] and because some information is better than no information. Of note, important caveats apply—such as the trial should be unbiased, reported properly, and published irrespective of the results, thereby becoming available for meta-analysis.[118] On the other hand, many medical researchers worry that underpowered trials with indeterminate results will remain unpublished and insist that all trials should individually have "sufficient power." This debate will continue, and members of the CONSORT Group have varying views. Critically however, the debate and those views are immaterial to reporting a trial. Whatever the power of a trial, authors need to properly report their intended size with all their methods and assumptions.[118] That transparently reveals the power of the trial to readers and gives them a measure by which to assess whether the trial attained its planned size.

In some trials, interim analyses are used to help decide whether to stop early or to continue recruiting sometimes beyond the planned trial end (see item 7b). If the actual sample size differed from the originally intended sample size for some other reason (for example, because of poor recruitment or revision of the target sample size), the explanation should be given.

Reports of studies with small samples frequently include the erroneous conclusion that the intervention groups do not differ, when in fact too few patients were studied to make such a claim.[120] Reviews of published trials have consistently found that a high proportion of trials have low power to detect clinically meaningful treatment effects.[121-123] In reality, small but clinically meaningful true differences are much more likely than large differences to exist, but large trials are required to detect them.[124]

In general, the reported sample sizes in trials seem small. The median sample size was 54 patients in 196 trials in arthritis,[108] 46 patients in 73 trials in dermatology,[8] and 65 patients in 2000 trials in schizophrenia.[33] These small sample sizes are consistent with those of a study of 519 trials indexed in PubMed in December 2000[16] and a similar cohort of trials (n=616) indexed in PubMed in 2006,[17] where the median number of patients recruited for parallel group trials was 80 across both years. Moreover, many reviews have found that few authors report how they determined the sample size.[8 14 32 33 123]

There is little merit in a post hoc calculation of statistical power using the results of a trial; the power is then appropriately indicated by confidence intervals (see item 17).[125]

Item 7b. When applicable, explanation of any interim analyses and stopping guidelines

Examples—"Two interim analyses were performed during the trial. The levels of significance maintained an overall P value of 0.05 and were calculated according to the O'Brien-Fleming stopping boundaries. This final analysis used a Z score of 1.985 with an associated P value of 0.0471."[126]

"An independent data and safety monitoring board periodically reviewed the efficacy and safety data. Stopping rules were based on modified Haybittle-Peto boundaries of 4 SD in the first half of the study and 3 SD in the second half for efficacy data, and 3 SD in the first half of the study and 2 SD in the second half for safety data. Two formal interim analyses of efficacy were performed when 50% and 75% of the expected number of primary events had accrued; no correction of the reported P value for these interim tests was performed."[127]

Explanation—Many trials recruit participants over a long period. If an intervention is working particularly well or badly, the study may need to be ended early for ethical reasons. This concern can be addressed by examining results as the data accumulate, preferably by an independent data monitoring

committee. However, performing multiple statistical examinations of accumulating data without appropriate correction can lead to erroneous results and interpretations.[128] If the accumulating data from a trial are examined at five interim analyses that use a P value of 0.05, the overall false positive rate is nearer to 19% than to the nominal 5%.

Several group sequential statistical methods are available to adjust for multiple analyses,[129-131] and their use should be pre-specified in the trial protocol. With these methods, data are compared at each interim analysis, and a P value less than the critical value specified by the group sequential method indicates statistical significance. Some trialists use group sequential methods as an aid to decision making,[132] whereas others treat them as a formal stopping rule (with the intention that the trial will cease if the observed P value is smaller than the critical value).

Authors should report whether they or a data monitoring committee took multiple "looks" at the data and, if so, how many there were, what triggered them, the statistical methods used (including any formal stopping rule), and whether they were planned before the start of the trial, before the data monitoring committee saw any interim data by allocation, or some time thereafter. This information is often not included in published trial reports,[133] even in trials that report stopping earlier than planned.[134]

Item 8a. Method used to generate the random allocation sequence

Examples—"Independent pharmacists dispensed either active or placebo inhalers according to a computer generated randomisation list."[63]

"For allocation of the participants, a computer-generated list of random numbers was used."[135]

Explanation—Participants should be assigned to comparison groups in the trial on the basis of a chance (random) process characterised by unpredictability (see box 1). Authors should provide sufficient information that the reader can assess the methods used to generate the random allocation sequence and the likelihood of bias in group assignment. It is important that information on the process of randomisation is included in the body of the main article and not as a separate supplementary file; where it can be missed by the reader.

The term "random" has a precise technical meaning. With random allocation, each participant has a known probability of receiving each intervention before one is assigned, but the assigned intervention is determined by a chance process and cannot be predicted. However, "random" is often used inappropriately in the literature to describe trials in which non-random, deterministic allocation methods were used, such as alternation, hospital numbers, or date of birth. When investigators use such non-random methods, they should describe them precisely and should not use the term "random" or any variation of it. Even the term "quasi-random" is unacceptable for describing such trials. Trials based on non-random methods generally yield biased results.[2-4 136] Bias presumably arises from the inability to conceal these allocation systems adequately (see item 9).

Many methods of sequence generation are adequate. However, readers cannot judge adequacy from such terms as "random allocation," "randomisation," or "random" without further elaboration. Authors should specify the method of sequence generation, such as a random-number table or a computerised random number generator. The sequence may be generated by the process of minimisation, a non-random but generally acceptable method (see box 2).

In some trials, participants are intentionally allocated in unequal numbers to each intervention: for example, to gain more experience with a new procedure or to limit costs of the trial. In such cases, authors should report the randomisation ratio (for example, 2:1 or two treatment participants per each control participant) (see item 3a).

In a representative sample of PubMed indexed trials in 2000, only 21% reported an adequate approach to random sequence generation[16]; this increased to 34% for a similar

Box 2 | Randomisation and minimisation

- **Simple randomisation—**Pure randomisation based on a single allocation ratio is known as simple randomisation. Simple randomisation with a 1:1 allocation ratio is analogous to a coin toss, although we do not advocate coin tossing for randomisation in an RCT. "Simple" is somewhat of a misnomer. While other randomisation schemes sound complex and more sophisticated, in reality, simple randomisation is elegantly sophisticated in that it is more unpredictable and surpasses the bias prevention levels of all other alternatives.
- **Restricted randomisation—**Any randomised approach that is not simple randomisation. Blocked randomisation is the most common form. Other means of restricted randomisation include replacement, biased coin, and urn randomisation, although these are used much less frequently.[141]
- **Blocked randomisation—**Blocking is used to ensure that comparison groups will be generated according to a predetermined ratio, usually 1:1 or groups of approximately the same size. Blocking can be used to ensure close balance of the numbers in each group at any time during the trial. For every block of eight participants, for example, four would be allocated to each arm of the trial.[142] Improved balance comes at the cost of reducing the unpredictability of the sequence. Although the order of interventions varies randomly within each block, a person running the trial could deduce some of the next treatment allocations if he or she knew the block size.[143] Blinding the interventions, using larger block sizes, and randomly varying the block size can ameliorate this problem.
- **Stratified randomisation—**Stratification is used to ensure good balance of participant characteristics in each group. By chance, particularly in small trials, study groups may not be well matched for baseline characteristics, such as age and stage of disease. This weakens the trial's credibility.[144] Such imbalances can be avoided without sacrificing the advantages of randomisation. Stratification ensures that the numbers of participants receiving each intervention are closely balanced within each stratum. Stratified randomisation is achieved by performing a separate randomisation procedure within each of two or more subsets of participants (for example, those defining each study centre, age, or disease severity). Stratification by centre is common in multicentre trials. Stratification requires some form of restriction (such as blocking within strata). Stratification without blocking is ineffective.
- **Minimisation—**Minimisation ensures balance between intervention groups for several selected patient factors (such as age).[22 60] The first patient is truly randomly allocated; for each subsequent participant, the treatment allocation that minimises the imbalance on the selected factors between groups at that time is identified. That allocation may then be used, or a choice may be made at random with a heavy weighting in favour of the intervention that would minimise imbalance (for example, with a probability of 0.8). The use of a random component is generally preferable. Minimisation has the advantage of making small groups closely similar in terms of participant characteristics at all stages of the trial. Minimisation offers the only acceptable alternative to randomisation, and some have argued that it is superior.[145] On the other hand, minimisation lacks the theoretical basis for eliminating bias on all known and unknown factors. Nevertheless, in general, trials that use minimisation are considered methodologically equivalent to randomised trials, even when a random element is not incorporated.

cohort of PubMed indexed trials in 2006.[17] In more than 90% of these cases, researchers used a random number generator on a computer or a random number table.

Item 8b. Type of randomisation; details of any restriction (such as blocking and block size)

Examples—"Randomization sequence was created using Stata 9.0 (StataCorp, College Station, TX) statistical software and was stratified by center with a 1:1 allocation using random block sizes of 2, 4, and 6."[137]

"Participants were randomly assigned following simple randomization procedures (computerized random numbers) to 1 of 2 treatment groups."[138]

Explanation—In trials of several hundred participants or more simple randomisation can usually be trusted to generate similar numbers in the two trial groups[139] and to generate groups that are roughly comparable in terms of known and unknown prognostic variables.[140] For smaller trials (see item 7a)—and even for trials that are not intended to be small, as they may stop before reaching their target size—some restricted randomisation (procedures to help achieve balance between groups in size or characteristics) may be useful (see box 2).

It is important to indicate whether no restriction was used, by stating such or by stating that "simple randomisation" was done. Otherwise, the methods used to restrict the randomisation, along with the method used for random selection, should be specified. For block randomisation, authors should provide details on how the blocks were generated (for example, by using a permuted block design with a computer random number generator), the block size or sizes, and whether the block size was fixed or randomly varied. If the trialists became aware of the block size(s), that information should also be reported as such knowledge could lead to code breaking. Authors should specify whether stratification was used, and if so, which factors were involved (such as recruitment site, sex, disease stage), the categorisation cut-off values within strata, and the method used for restriction. Although stratification is a useful technique, especially for smaller trials, it is complicated to implement and may be impossible if many stratifying factors are used. If minimisation (see box 2) was used, it should be explicitly identified, as should the variables incorporated into the scheme. If used, a random element should be indicated.

Only 9% of 206 reports of trials in specialty journals[23] and 39% of 80 trials in general medical journals reported use of stratification.[32] In each case, only about half of the reports mentioned the use of restricted randomisation. However, these studies and that of Adetugbo and Williams[8] found that the sizes of the treatment groups in many trials were the same or quite similar, yet blocking or stratification had not been mentioned. One possible explanation for the close balance in numbers is underreporting of the use of restricted randomisation.

Item 9. Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

Examples—"The doxycycline and placebo were in capsule form and identical in appearance. They were prepacked in bottles and consecutively numbered for each woman according to the randomisation schedule. Each woman was assigned an order number and received the capsules in the corresponding prepacked bottle."[146]

"The allocation sequence was concealed from the researcher (JR) enrolling and assessing participants in sequentially numbered, opaque, sealed and stapled envelopes. Aluminium foil inside the envelope was used to render the envelope impermeable to intense light. To prevent subversion of the allocation sequence, the name and date of birth of the participant was written on the envelope and a video tape made of the sealed envelope with participant details visible. Carbon paper inside the envelope transferred the information onto the allocation card inside the envelope and a second researcher (CC) later viewed video tapes to ensure envelopes were still sealed when participants' names were written on them. Corresponding envelopes were opened only after the enrolled participants completed all baseline assessments and it was time to allocate the intervention."[147]

Explanation—Item 8a discussed generation of an unpredictable sequence of assignments. Of considerable importance is how this sequence is applied when participants are enrolled into the trial (see box 1). A generated allocation schedule should be implemented by using allocation concealment,[23] a critical mechanism that prevents foreknowledge of treatment assignment and thus shields those who enroll participants from being influenced by this knowledge. The decision to accept or reject a participant should be made, and informed consent should be obtained from the participant, in ignorance of the next assignment in the sequence.[148]

The allocation concealment should not be confused with blinding (see item 11). Allocation concealment seeks to prevent selection bias, protects the assignment sequence until allocation, and can always be successfully implemented.[2] In contrast, blinding seeks to prevent performance and ascertainment bias, protects the sequence after allocation, and cannot always be implemented.[23] Without adequate allocation concealment, however, even random, unpredictable assignment sequences can be subverted.[2 149]

Centralised or "third-party" assignment is especially desirable. Many good allocation concealment mechanisms incorporate external involvement. Use of a pharmacy or central telephone randomisation system are two common techniques. Automated assignment systems are likely to become more common.[150] When external involvement is not feasible, an excellent method of allocation concealment is the use of numbered containers. The interventions (often drugs) are sealed in sequentially numbered identical containers according to the allocation sequence.[151] Enclosing assignments in sequentially numbered, opaque, sealed envelopes can be a good allocation concealment mechanism if it is developed and monitored diligently. This method can be corrupted, however, particularly if it is poorly executed. Investigators should ensure that the envelopes are opaque when held to the light, and opened sequentially and only after the participant's name and other details are written on the appropriate envelope.[143]

A number of methodological studies provide empirical evidence to support these precautions.[152 153] Trials in which the allocation sequence had been inadequately or unclearly

Box 3 | Steps in a typical randomisation process

Sequence generation
- Generate allocation sequence by some random procedure

Allocation concealment
- Develop allocation concealment mechanism (such as numbered, identical bottles or sequentially numbered, sealed, opaque envelopes)
- Prepare the allocation concealment mechanism using the allocation sequence from the sequence generation step

Implementation
- Enrol participants:
  - Assess eligibility
  - Discuss the trial
  - Obtain informed consent
  - Enrol participant in trial
- Ascertain intervention assignment (such as opening next envelope)
- Administer intervention

concealed yielded larger estimates of treatment effects than did trials in which authors reported adequate allocation concealment. These findings provide strong empirical evidence that inadequate allocation concealment contributes to bias in estimating treatment effects.

Despite the importance of the mechanism of allocation concealment, published reports often omit such details. The mechanism used to allocate interventions was omitted in reports of 89% of trials in rheumatoid arthritis,[108] 48% of trials in obstetrics and gynaecology journals,[23] and 44% of trials in general medical journals.[32] In a more broadly representative sample of all randomised trials indexed on PubMed, only 18% reported any allocation concealment mechanism, but some of those reported mechanisms were inadequate.[16]

Item 10. Who generated the allocation sequence, who enrolled participants, and who assigned participants to interventions

Examples—"Determination of whether a patient would be treated by streptomycin and bed-rest (S case) or by bed-rest alone (C case) was made by reference to a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill; the details of the series were unknown to any of the investigators or to the co-ordinator ... After acceptance of a patient by the panel, and before admission to the streptomycin centre, the appropriate numbered envelope was opened at the central office; the card inside told if the patient was to be an S or a C case, and this information was then given to the medical officer of the centre."[24]

"Details of the allocated group were given on coloured cards contained in sequentially numbered, opaque, sealed envelopes. These were prepared at the NPEU and kept in an agreed location on each ward. Randomisation took place at the end of the 2nd stage of labour when the midwife considered a vaginal birth was imminent. To enter a women into the study, the midwife opened the next consecutively numbered envelope."[154]

"Block randomisation was by a computer generated random number list prepared by an investigator with no clinical involvement in the trial. We stratified by admission for an oncology related procedure. After the research nurse had obtained the patient's consent, she telephoned a contact who was independent of the recruitment process for allocation consignment."[155]

Explanation—As noted in item 9, concealment of the allocated intervention at the time of enrolment is especially important. Thus, in addition to knowing the methods used, it is also important to understand how the random sequence was implemented—specifically, who generated the allocation sequence, who enrolled participants, and who assigned participants to trial groups.

The process of randomising participants into a trial has three different steps: sequence generation, allocation concealment, and implementation (see box 3). Although the same people may carry out more than one process under each heading, investigators should strive for complete separation of the people involved with generation and allocation concealment from the people involved in the implementation of assignments. Thus, if someone is involved in the sequence generation or allocation concealment steps, ideally they should not be involved in the implementation step.

Even with flawless sequence generation and allocation concealment, failure to separate creation and concealment of the allocation sequence from assignment to study group may introduce bias. For example, the person who generated an allocation sequence could retain a copy and consult it when interviewing potential participants for a trial. Thus, that person could bias the enrolment or assignment process, regardless of the unpredictability of the assignment sequence. Investigators must then ensure that the assignment schedule is unpredictable and locked away (such as in a safe deposit box in a building rather inaccessible to the enrolment location) from even the person who generated it. The report of the trial should specify where the investigators stored the allocation list.

Item 11a. If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how

Examples—"Whereas patients and physicians allocated to the intervention group were aware of the allocated arm, outcome assessors and data analysts were kept blinded to the allocation."[156]

"Blinding and equipoise were strictly maintained by emphasising to intervention staff and participants that each diet adheres to healthy principles, and each is advocated by certain experts to be superior for long-term weight-loss. Except for the interventionists (dieticians and behavioural psychologists), investigators and staff were kept blind to diet assignment of the participants. The trial adhered to established procedures to maintain separation between staff that take outcome measurements and staff that deliver the intervention. Staff members who obtained outcome measurements were not informed of the diet group assignment. Intervention staff, dieticians and behavioural psychologists who delivered the intervention did not take outcome measurements. All investigators, staff, and participants were kept masked to outcome measurements and trial results."[157]

Explanation—The term "blinding" or "masking" refers to withholding information about the assigned interventions from people involved in the trial who may potentially be influenced by this knowledge. Blinding is an important safeguard against bias, particularly when assessing subjective outcomes.[153]

Benjamin Franklin has been credited as being the first to

use blinding in a scientific experiment.[158] He blindfolded participants so they would not know when he was applying mesmerism (a popular "healing fluid" of the 18th century) and in so doing showed that mesmerism was a sham. Based on this experiment, the scientific community recognised the power of blinding to reduce bias, and it has remained a commonly used strategy in scientific experiments.

Box 4, on blinding terminology, defines the groups of individuals (that is, participants, healthcare providers, data collectors, outcome adjudicators, and data analysts) who can potentially introduce bias into a trial through knowledge of the treatment assignments. Participants may respond differently if they are aware of their treatment assignment (such as responding more favourably when they receive the new treatment).[153] Lack of blinding may also influence compliance with the intervention, use of co-interventions, and risk of dropping out of the trial.

Unblinded healthcare providers may introduce similar biases, and unblinded data collectors may differentially assess outcomes (such as frequency or timing), repeat measurements of abnormal findings, or provide encouragement during performance testing. Unblinded outcome adjudicators may differentially assess subjective outcomes, and unblinded data analysts may introduce bias through the choice of analytical strategies, such as the selection of favourable time points or outcomes, and by decisions to remove patients from the analyses. These biases have been well documented.[71 153 159-162]

Blinding, unlike allocation concealment (see item 10), may not always be appropriate or possible. An example is a trial comparing levels of pain associated with sampling blood from the ear or thumb.[163] Blinding is particularly important when outcome measures involve some subjectivity, such as assessment of pain. Blinding of data collectors and outcome adjudicators is unlikely to matter for objective outcomes, such as death from any cause. Even then, however, lack of participant or healthcare provider blinding can lead to other problems, such as differential attrition.[164] In certain trials, especially surgical trials, blinding of participants and surgeons is often difficult or impossible, but blinding of data collectors and outcome adjudicators is often achievable. For example, lesions can be photographed before and after treatment and assessed by an external observer.[165] Regardless of whether blinding is possible, authors can and should always state who was blinded (that is, participants, healthcare providers, data collectors, and outcome adjudicators).

Unfortunately, authors often do not report whether blinding was used.[166] For example, reports of 51% of 506 trials in cystic fibrosis,[167] 33% of 196 trials in rheumatoid arthritis,[108] and 38% of 68 trials in dermatology[8] did not state whether blinding was used. Until authors of trials improve their reporting of blinding, readers will have difficulty in judging the validity of the trials that they may wish to use to guide their clinical practice.

The term masking is sometimes used in preference to blinding to avoid confusion with the medical condition of being without sight. However, "blinding" in its methodological sense seems to be understood worldwide and is acceptable for reporting clinical trials.[165 168]

Item 11b. If relevant, description of the similarity of interventions

Example—"Jamieson Laboratories Inc provided 500-mg immediate release niacin in a white, oblong, bisect caplet. We independently confirmed caplet content using high performance liquid chromatography... The placebo was matched to the study drug for taste, color, and size, and contained microcrystalline cellulose, silicon dioxide, dicalcium phosphate, magnesium stearate, and stearic acid."[172]

Explanation—Just as we seek evidence of concealment to assure us that assignment was truly random, we seek evidence of the method of blinding. In trials with blinding of participants or healthcare providers, authors should state the similarity of the characteristics of the interventions (such as appearance, taste, smell, and method of administration).[35 173]

Some people have advocated testing for blinding by

Box 4 | Blinding terminology

In order for a technical term to have utility it must have consistency in its use and interpretation. Authors of trials commonly use the term "double blind" and, less commonly, the terms "single blind"or "triple blind." A problem with this lexicon is that there is great variability in clinician interpretations and epidemiological textbook definitions of these terms.[169] Moreover, a study of 200 RCTs reported as double blind found 18 different combinations of groups actually blinded when the authors of these trials were surveyed, and about one in every five of these trials—reported as double blind—did not blind participants, healthcare providers, or data collectors.[170]

This research shows that terms are ambiguous and, as such, authors and editors should abandon their use. Authors should instead explicitly report the blinding status of the people involved for whom blinding may influence the validity of a trial.

Healthcare providers include all personnel (for example, physicians, chiropractors, physiotherapists, nurses) who care for the participants during the trial. Data collectors are the individuals who collect data on the trial outcomes. Outcome adjudicators are the individuals who determine whether a participant did experience the outcomes of interest.

Some researchers have also advocated blinding and reporting the blinding status of the data monitoring committee and the manuscript writers.[160] Blinding of these groups is uncommon, and the value of blinding them is debated.[171]

Sometimes one group of individuals (such as the healthcare providers) are the same individuals fulfilling another role in a trial (such as data collectors). Even if this is the case, the authors should explicitly state the blinding status of these groups to allow readers to judge the validity of the trial.

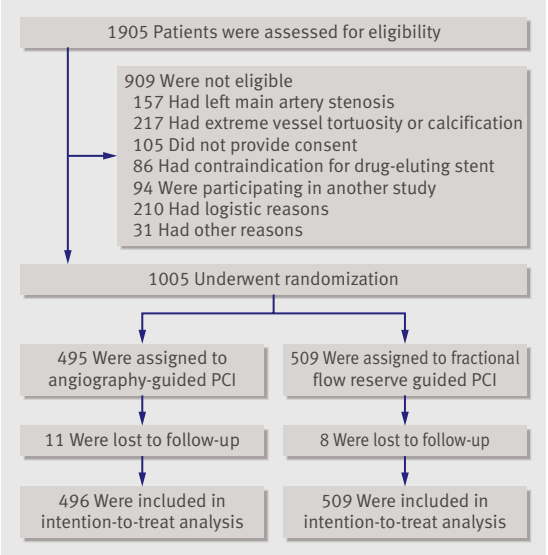


Fig 2 | Flow diagram of a multicentre trial of fractional flow reserve versus angiography for guiding percutaneous coronary intervention (PCI) (adapted from Tonino et al[313]). The diagram includes detailed information on the excluded participants.

asking participants or healthcare providers at the end of a trial whether they think the participant received the experimental or control intervention.[174] Because participants and healthcare providers will usually know whether the participant has experienced the primary outcome, this makes it difficult to determine if their responses reflect failure of blinding or accurate assumptions about the efficacy of the intervention.[175] Given the uncertainty this type of information provides, we have removed advocating reporting this type of testing for blinding from the CONSORT 2010 Statement. We do, however, advocate that the authors report any known compromises in blinding. For example, authors should report if it was necessary to unblind any participants at any point during the conduct of a trial.

Item 12a. Statistical methods used to compare groups for primary and secondary outcomes

Example—"The primary endpoint was change in bodyweight during the 20 weeks of the study in the intention-to-treat population ... Secondary efficacy endpoints included change in waist circumference, systolic and diastolic blood pressure, prevalence of metabolic syndrome ... We used an analysis of covariance (ANCOVA) for the primary endpoint and for secondary endpoints waist circumference, blood pressure, and patient-reported outcome scores; this was supplemented by a repeated measures analysis. The ANCOVA model included treatment, country, and sex as fixed effects, and bodyweight at randomisation as covariate. We aimed to assess whether data provided evidence of superiority of each liraglutide dose to placebo (primary objective) and to orlistat (secondary objective)."[176]

Explanation—Data can be analysed in many ways, some of which may not be strictly appropriate in a particular situation. It is essential to specify which statistical procedure was used for each analysis, and further clarification may be necessary in the results section of the report. The principle to follow is to, "Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results" (www.icmje.org). It is also important to describe details of the statistical analysis such as intention-to-treat analysis (see box 6).

Almost all methods of analysis yield an estimate of the treatment effect, which is a contrast between the outcomes in the comparison groups. Authors should accompany this by a confidence interval for the estimated effect, which indicates a central range of uncertainty for the true treatment effect. The confidence interval may be interpreted as the range of values for the treatment effect that is compatible with the observed data. It is customary to present a 95% confidence interval, which gives the range expected to include the true value in 95 of 100 similar studies.

Study findings can also be assessed in terms of their statistical significance. The P value represents the probability that the observed data (or a more extreme result) could have arisen by chance when the interventions did not truly differ. Actual P values (for example, P=0.003) are strongly preferable to imprecise threshold reports such as P<0.05.[48 177]

Standard methods of analysis assume that the data are "independent." For controlled trials, this usually means that there is one observation per participant. Treating multiple observations from one participant as independent data is a serious error; such data are produced when outcomes can be measured on different parts of the body, as in dentistry or rheumatology. Data analysis should be based on counting each participant once[178 179] or should be done by using more complex statistical procedures.[180] Incorrect analysis of multiple observations per individual was seen in 123 (63%) of 196 trials in rheumatoid arthritis.[108]

Item 12b. Methods for additional analyses, such as subgroup analyses and adjusted analyses

Examples—"Proportions of patients responding were compared between treatment groups with the Mantel-Haenszel χ^2 test, adjusted for the stratification variable, methotrexate use."[103]

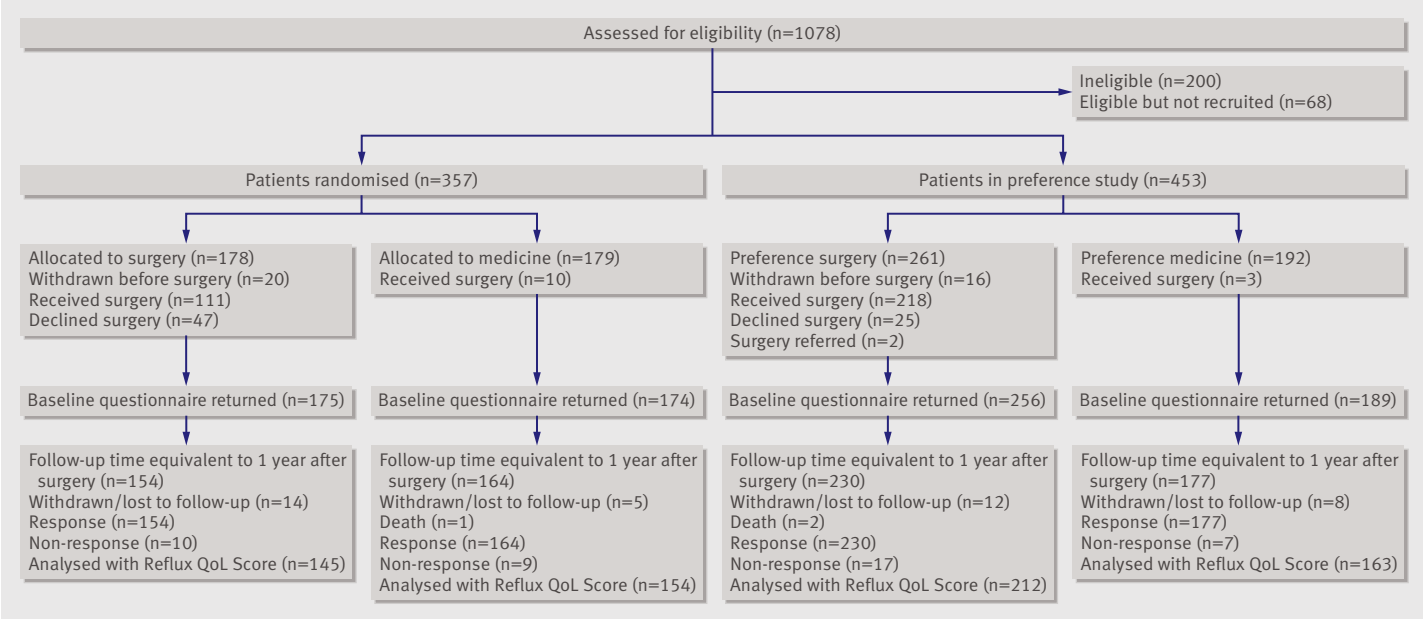


Fig 3 | Flow diagram of minimal surgery compared with medical management for chronic gastro-oesophageal reflux disease (adapted from Grant et al[196]). The diagram shows a multicentre trial with a parallel non-randomised preference group.

Table 3 | Information required to document the flow of participants through each stage of a randomised trial

Stage	Number of people included	Number of people not included or excluded	Rationale
Enrolment	People evaluated for potential enrolment	People who did not meet the inclusion criteria or met the inclusion criteria but declined to be enrolled	These counts indicate whether trial participants were likely to be representative of all patients seen; they are relevant to assessment of external validity only, and they are often not available.
Randomisation	Participants randomly assigned		Crucial count for defining trial size and assessing whether a trial has been analysed by intention to treat
Treatment allocation	Participants who completed treatment as allocated, by study group	Participants who did not complete treatment as allocated, by study group	Important counts for assessment of internal validity and interpretation of results; reasons for not receiving treatment as allocated should be given.
Follow-up	Participants who completed treatment as allocated, by study group	Participants who did not complete treatment as allocated, by study group	Important counts for assessment of internal validity and interpretation of results; reasons for not completing treatment or follow-up should be given.
	Participants who completed follow-up as planned, by study group	Participants who did not complete follow-up as planned, by study group	
Analysis	Participants included in main analysis, by study group	Participants excluded from main analysis, by study group	Crucial count for assessing whether a trial has been analysed by intention to treat; reasons for excluding participants should be given.

"Pre-specified subgroup analyses according to antioxidant treatment assignment(s), presence or absence of prior CVD, dietary folic acid intake, smoking, diabetes, aspirin, hormone therapy, and multivitamin use were performed using stratified Cox proportional hazards models. These analyses used baseline exposure assessments and were restricted to participants with nonmissing subgroup data at baseline."[181]

Explanation—As is the case for primary analyses, the method of subgroup analysis should be clearly specified. The strongest analyses are those that look for evidence of a difference in treatment effect in complementary subgroups (for example, older and younger participants), a comparison known as a test of interaction.[182 183] A common but misleading approach is to compare P values for separate analyses of the treatment effect in each group. It is incorrect to infer a subgroup effect (interaction) from one significant and one non-significant P value.[184] Such inferences have a high false positive rate.

Because of the high risk for spurious findings, subgroup analyses are often discouraged.[14 185] Post hoc subgroup comparisons (analyses done after looking at the data) are especially likely not to be confirmed by further studies. Such analyses do not have great credibility.

In some studies, imbalances in participant characteristics are adjusted for by using some form of multiple regression analysis. Although the need for adjustment is much less in RCTs than in epidemiological studies, an adjusted analysis may be sensible, especially if one or more variables is thought to be prognostic.[186] Ideally, adjusted analyses should be specified in the study protocol (see item 24). For example, adjustment is often recommended for any stratification variables (see item 8b) on the principle that the analysis strategy should follow the design. In RCTs, the decision to adjust should not be determined by whether baseline differences are statistically significant (see item 16).[183 187] The rationale for any adjusted analyses and the statistical methods used should be specified.

Authors should clarify the choice of variables that were adjusted for, indicate how continuous variables were handled, and specify whether the analysis was planned or suggested by the data.[188] Reviews of published studies show that reporting of adjusted analyses is inadequate with regard to all of these aspects.[188-191]

Results

Item 13. Participant flow (a diagram is strongly recommended)

Item 13a. For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome

Examples—See figs 2 and 3.

Explanation—The design and conduct of some RCTs is straightforward, and the flow of participants, particularly were there are no losses to follow-up or exclusions, through each phase of the study can be described adequately in a few sentences. In more complex studies, it may be difficult for readers to discern whether and why some participants did not receive the treatment as allocated, were lost to follow-up, or were excluded from the analysis.[51] This information is crucial for several reasons. Participants who were excluded after allocation are unlikely to be representative of all participants in the study. For example, patients may not be available for follow-up evaluation because they experienced an acute exacerbation of their illness or harms of treatment.[22 192]

Attrition as a result of loss to follow up, which is often unavoidable, needs to be distinguished from investigator-determined exclusion for such reasons as ineligibility, withdrawal from treatment, and poor adherence to the trial protocol. Erroneous conclusions can be reached if participants are excluded from analysis, and imbalances in such omissions between groups may be especially indicative of bias.[192-194] Information about whether the investigators included in the analysis all participants who underwent randomisation, in the groups to which they were originally allocated (intention-to-treat analysis (see item 16 and box 6)), is therefore of particular importance. Knowing the number of participants who did not receive the intervention as allocated or did not complete treatment permits the reader to assess to what extent the estimated efficacy of therapy might be underestimated in comparison with ideal circumstances.

If available, the number of people assessed for eligibility should also be reported. Although this number is relevant to external validity only and is arguably less important than the other counts,[195] it is a useful indicator of whether trial participants were likely to be representative of all eligible participants.

A review of RCTs published in five leading general and internal medicine journals in 1998 found that reporting of the

Box 5 | Early stopping

RCTs can end when they reach their sample size goal, their event count goal, their length of follow-up goal, or when they reach their scheduled date of closure. In these situations the trial will stop in a manner independent of its results, and stopping is unlikely to introduce bias in the results. Alternatively, RCTs can stop earlier than planned because of the result of an interim analysis showing larger than expected benefit or harm on the experimental intervention. Also RCTs can stop earlier than planned when investigators find evidence of no important difference between experimental and control interventions (that is, stopping for futility). In addition, trials may stop early because the trial becomes unviable: funding vanishes, researchers cannot access eligible patients or study interventions, or the results of other studies make the research question irrelevant.

Full reporting of why a trial ended is important for evidence based decision making (see item 14b). Researchers examining why 143 trials stopped early for benefit found that many failed to report key methodological information regarding how the decision to stop was reached—the planned sample size (n=28), interim analysis after which the trial was stopped (n=45), or whether a stopping rule informed the decision (n=48).[134] Item 7b of the checklist requires the reporting of timing of interim analyses, what triggered them, how many took place, whether these were planned or ad hoc, and whether there were statistical guidelines and stopping rules in place a priori. Furthermore, it is helpful to know whether an independent data monitoring committee participated in the analyses (and who composed it, with particular attention to the role of the funding source) and who made the decision to stop. Often the data safety and monitoring committee makes recommendations and the funders (sponsors) or the investigators make the decision to stop.

Trials that stop early for reasons apparently independent of trial findings, and trials that reach their planned termination, are unlikely to introduce bias by stopping.[207] In these cases, the authors should report whether interim analyses took place and whether these results were available to the funder.

The push for trials that change the intervention in response to interim results, thus enabling a faster evaluation of promising interventions for rapidly evolving and fatal conditions, will require even more careful reporting of the process and decision to stop trials early.[208]

flow of participants was often incomplete, particularly with regard to the number of participants receiving the allocated intervention and the number lost to follow-up.[51] Even information as basic as the number of participants who underwent randomisation and the number excluded from analyses was not available in up to 20% of articles.[51] Reporting was considerably more thorough in articles that included a diagram of the flow of participants through a trial, as recommended by CONSORT. This study informed the design of the revised flow diagram in the revised CONSORT statement.[52-54] The suggested template is shown in fig 1, and the counts required are described in detail in table 3.

Some information, such as the number of individuals assessed for eligibility, may not always be known,[14] and, depending on the nature of a trial, some counts may be more relevant than others. It will sometimes be useful or necessary to adapt the structure of the flow diagram to a particular trial. In some situations, other information may usefully be added. For example, the flow diagram of a parallel group trial of minimal surgery compared with medical management for chronic gastro-oesophageal reflux also included a parallel non-randomised preference group (see fig 3).[196]

The exact form and content of the flow diagram may be varied according to specific features of a trial. For example, many trials of surgery or vaccination do not include the possibility of discontinuation. Although CONSORT strongly recommends using this graphical device to communicate participant flow throughout the study, there is no specific, prescribed format.

Item 13b. For each group, losses and exclusions after randomisation, together with reasons

Examples—"There was only one protocol deviation, in a woman in the study group. She had an abnormal pelvic measurement and was scheduled for elective caesarean section. However, the attending obstetrician judged a trial of labour acceptable; caesarean section was done when there was no progress in the first stage of labour."[197]

"The monitoring led to withdrawal of nine centres, in which existence of some patients could not be proved, or other serious violations of good clinical practice had occurred."[198]

Explanation—Some protocol deviations may be reported in

Table 4 | Example of reporting baseline demographic and clinical characteristics.* (Adapted from table 1 of Yusuf et al[209])

	Telmisartan (N=2954)	Placebo (N=2972)
Age (years)	66.9 (7.3)	66.9 (7.4)
Sex (female)	1280 (43.3%)	1267 (42.6%)
Smoking status:		
Current	293 (9.9%)	289 (9.7%)
Past	1273 (43.1%)	1283 (43.2%)
Ethnic origin:		
Asian	637 (21.6%)	624 (21.0%)
Arab	37 (1.3%)	40 (1.3%)
African	51 (1.7%)	55 (1.9%)
European	1801 (61.0%)	1820 (61.2%)
Native or Aboriginal	390 (13.2%)	393 (13.2%)
Other	38 (1.3%)	40 (1.3%)
Blood pressure (mm Hg)	140.7 (16.8/81.8) (10.1)	141.3 (16.4/82.0) (10.2)
Heart rate (beats per min)	68.8 (11.5)	68.8 (12.1)
Cholesterol (mmol/l):		
Total	5.09 (1.18)	5.08 (1.15)
LDL	3.02 (1.01)	3.03 (1.02)
HDL	1.27 (0.37)	1.28 (0.41)
Coronary artery disease	2211 (74.8%)	2207 (74.3%)
Myocardial infarction	1381 (46.8%)	1360 (45.8%)
Angina pectoris	1412 (47.8%)	1412 (47.5%)
Peripheral artery disease	349 (11.8%)	323 (10.9%)
Hypertension	2259 (76.5%)	2269 (76.3%)
Diabetes	1059 (35.8%)	1059 (35.6%)

*Data are means (SD) or numbers (%).

the flow diagram (see item 13a)—for example, participants who did not receive the intended intervention. If participants were excluded after randomisation (contrary to the intention-to-treat principle) because they were found not to meet eligibility criteria (see item 16), they should be included in the flow diagram. Use of the term "protocol deviation" in published articles is not sufficient to justify exclusion of participants after randomisation. The nature of the protocol deviation and the exact reason for excluding participants after randomisation should always be reported.

Item 14a. Dates defining the periods of recruitment and follow-up

Example—"Age-eligible participants were recruited ... from February 1993 to September 1994 ... Participants attended clinic visits at the time of randomisation (baseline) and at 6-month intervals for 3 years."[199]

Explanation—Knowing when a study took place and over what period participants were recruited places the study in historical context. Medical and surgical therapies, including concurrent therapies, evolve continuously and may affect the routine care given to participants during a trial. Knowing the rate at which participants were recruited may also be useful, especially to other investigators.

The length of follow-up is not always a fixed period after randomisation. In many RCTs in which the outcome is time to an event, follow-up of all participants is ended on a specific date. This date should be given, and it is also useful to report the minimum, maximum, and median duration of follow-up.[200 201]

A review of reports in oncology journals that used survival analysis, most of which were not RCTs, [201] found that nearly 80% (104 of 132 reports) included the starting and ending dates for accrual of patients, but only 24% (32 of 132 reports) also reported the date on which follow-up ended.

Item 14b. Why the trial ended or was stopped

Examples—"At the time of the interim analysis, the total follow-up included an estimated 63% of the total number of patient-years that would have been collected at the end of the study, leading to a threshold value of 0.0095, as determined by the Lan-DeMets alpha-spending function method ... At the interim analysis, the RR was 0.37 in the intervention group, as compared with the control group, with a p value of 0.00073, below the threshold value. The Data and Safety Monitoring Board advised the investigators to interrupt the trial and offer circumcision to the control group, who were then asked to come to the investigation centre, where MC (medical circumcision) was advised and proposed ... Because the study was interrupted, some participants did not have a full follow-up on that date, and their visits that were not yet completed are described as "planned" in this article."[202]

"In January 2000, problems with vaccine supply necessitated the temporary nationwide replacement of the whole cell component of the combined DPT/Hib vaccine with acellular pertussis vaccine. As this vaccine has a different local reactogenicity profile, we decided to stop the trial early."[203]

Explanation—Arguably, trialists who arbitrarily conduct unplanned interim analyses after very few events accrue using no statistical guidelines run a high risk of "catching" the data at a random extreme, which likely represents a large overestimate of treatment benefit.[204]

Readers will likely draw weaker inferences from a trial that was truncated in a data-driven manner versus one that reports its findings after reaching a goal independent of results. Thus, RCTs should indicate why the trial came to an end (see box 5). The report should also disclose factors extrinsic to the trial that affected the decision to stop the trial, and who made the decision to stop the trial, including reporting the role the funding agency played in the deliberations and in the decision to stop the trial.[134]

A systematic review of 143 RCTs stopped earlier than planned for benefit found that these trials reported stopping after accruing a median of 66 events, estimated a median relative risk of 0.47 and a strong relation between the number of events accrued and the size of the effect, with smaller trials with fewer events yielding the largest treatment effects (odds ratio 31, 95% confidence interval 12 to 82).[134] While an increasing number of trials published in high impact medical journals report stopping early, only 0.1% of trials reported stopping early for benefit, which contrasts with estimates arising from simulation studies[205] and surveys of data safety and monitoring committees.[206] Thus, many trials accruing few participants and reporting large treatment effects may have been stopped earlier than planned but failed to report this action.

Item 15. A table showing baseline demographic and clinical characteristics for each group

Example—See table 4

Explanation—Although the eligibility criteria (see item 4a) indicate who was eligible for the trial, it is also important to know the characteristics of the participants who were actually included. This information allows readers, especially clinicians, to judge how relevant the results of a trial might be to an individual patient.

Randomised trials aim to compare groups of participants that differ only with respect to the intervention (treatment). Although proper random assignment prevents selection bias, it does not guarantee that the groups are equivalent at baseline. Any differences in baseline characteristics are, however, the result of chance rather than bias.[32] The study groups should be compared at baseline for important demographic and clinical characteristics so that readers can assess how similar they were. Baseline data are especially valuable for outcomes that can also be measured at the start of the trial (such as blood pressure).

Baseline information is most efficiently presented in a table (see table 4). For continuous variables, such as weight or blood pressure, the variability of the data should be reported, along with average values. Continuous variables can be summarised for each group by the mean and standard deviation. When continuous data have an asymmetrical distribution, a preferable approach may be to quote the median and a centile range (such as the 25th and 75th centiles).[177] Standard errors and confidence intervals are not appropriate for describing variability—they are inferential rather than descriptive statistics. Variables with a small number of ordered categories (such as stages of dis-

Box 6 | Intention-to-treat analysis

The special strength of the RCT is the avoidance of bias when allocating interventions to trial participants (see box 1). That strength allows strong inferences about cause and effect that are not justified with other study designs. In order to preserve fully the huge benefit of randomisation we should include all randomised participants in the analysis, all retained in the group to which they were allocated. Those two conditions define an "intention-to-treat" analysis, which is widely recommended as the preferred analysis strategy.[18 223] Intention-to-treat analysis corresponds to analysing the groups exactly as randomised. Strict intention-to-treat analysis is often hard to achieve for two main reasons—missing outcomes for some participants and non-adherence to the trial protocol.

Missing outcomes

Many trialists exclude patients without an observed outcome. Often this is reasonable, but once any randomised participants are excluded the analysis is not strictly an intention-to-treat analysis. Indeed, most randomised trials have some missing observations. Trialists effectively must choose between omitting the participants without final outcome data or imputing their missing outcome data.[225] A "complete case" (or "available case") analysis includes only those whose outcome is known. While a few missing outcomes will not cause a problem, in half of trials more than 10% of randomised patients may have missing outcomes.[226] This common approach will lose power by reducing the sample size, and bias may well be introduced if being lost to follow-up is related to a patient's response to treatment. There should be concern when the frequency or the causes of dropping out differ between the intervention groups.

Participants with missing outcomes can be included in the analysis only if their outcomes are imputed (that is, their outcomes are estimated from other information that was collected). Imputation of the missing data allows the analysis to conform to intention-to-treat analysis but requires strong assumptions, which may be hard to justify.[227] Simple imputation methods are appealing, but their use may be inadvisable. In particular, a widely used method is "last observation carried forward" in which missing final values of the outcome variable are replaced by the last known value before the participant was lost to follow up. This is appealing through its simplicity, but the method may introduce bias,[228] and no allowance is made for the uncertainty of imputation.[229] Many authors have severely criticised last observation carried forward.[229-231]

Non-adherence to the protocol

A separate issue is that the trial protocol may not have been followed fully for some trial participants. Common examples are participants who did not meet the inclusion criteria (such as wrong diagnosis, too young), received a proscribed co-intervention, did not take all the intended treatment, or received a different treatment or no intervention. The simple way to deal with any protocol deviations is to ignore them: all participants can be included in the analysis regardless of adherence to the protocol, and this is the intention-to-treat approach. Thus, exclusion of any participants for such reasons is incompatible with intention-to-treat analysis.

The term "modified intention-to-treat" is quite widely used to describe an analysis that excludes participants who did not adequately adhere to the protocol, in particular those who did not receive a defined minimum amount of the intervention.[232] An alternative term is "per protocol." Though a per protocol analysis may be appropriate in some settings, it should be properly labelled as a non-randomised, observational comparison. Any exclusion of patients from the analysis compromises the randomisation and may lead to bias in the results.

Like "intention-to-treat," none of these other labels reliably clarifies exactly which patients were included. Thus, in the CONSORT checklist we have dropped the specific request for intention-to-treat analysis in favour of a clear description of exactly who was included in each analysis.

ease I to IV) should not be treated as continuous variables; instead, numbers and proportions should be reported for each category.[48 177]

Unfortunately significance tests of baseline differences are still common[23 32 210]; they were reported in half of 50 RCTs trials published in leading general journals in 1997.[183] Such significance tests assess the probability that observed baseline differences could have occurred by chance; however, we already know that any differences are caused by chance. Tests of baseline differences are not necessarily wrong, just illogical.[211] Such hypothesis testing is superfluous and can mislead investigators and their readers. Rather, comparisons at baseline should be based on consideration of the prognostic strength of the variables measured and the size of any chance imbalances that have occurred.[211]

Table 5 | Example of reporting of summary results for each study group (binary outcomes).* (Adapted from table 2 of Mease et al[103])

Endpoint	Number (%)		Risk difference (95% CI)
	Etanercept (n=30)	Placebo (n=30)	
Primary endpoint			
Achieved PsARC at 12 weeks	26 (87)	7 (23)	63% (44 to 83)
Secondary endpoint			
Proportion of patients meeting ACR criteria:			
ACR20	22 (73)	4 (13)	60% (40 to 80)
ACR50	15 (50)	1 (3)	47% (28 to 66)
ACR70	4 (13)	0 (0)	13% (1 to 26)

*See also example for item 6a.
PsARC=psoriatic arthritis response criteria. ACR=American College of Rheumatology.

Table 6 | Example of reporting of summary results for each study group (continuous outcomes). (Adapted from table 3 of van Linschoten[234])

	Exercise therapy (n=65)		Control (n=66)		Adjusted difference* (95% CI) at 12 months
	Baseline (mean (SD))	12 months (mean (SD))	Baseline (mean (SD))	12 months (mean (SD))	
Function score (0-100)	64.4 (13.9)	83.2 (14.8)	65.9 (15.2)	79.8 (17.5)	4.52 (−0.73 to 9.76)
Pain at rest (0-100)	4.14 (2.3)	1.43 (2.2)	4.03 (2.3)	2.61 (2.9)	−1.29 (−2.16 to −0.42)
Pain on activity (0-100)	6.32 (2.2)	2.57 (2.9)	5.97 (2.3)	3.54 (3.38)	−1.19 (−2.22 to −0.16)

*Function score adjusted for baseline, age, and duration of symptoms.

Item 16. For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups

Examples—"The primary analysis was intention-to-treat and involved all patients who were randomly assigned."[212]

"One patient in the alendronate group was lost to follow up; thus data from 31 patients were available for the intention-to-treat analysis. Five patients were considered protocol violators ... consequently 26 patients remained for the per-protocol analyses."[213]

Explanation—The number of participants in each group is an essential element of the analyses. Although the flow diagram (see item 13a) may indicate the numbers of participants analysed, these numbers often vary for different outcome measures. The number of participants per group should be given for all analyses. For binary outcomes, (such as risk ratio and risk difference) the denominators or event rates should also be reported. Expressing results as fractions also aids the reader in assessing whether some of the randomly assigned participants were excluded from the analysis. It follows that results should not be presented solely as summary measures, such as relative risks.

Table 7 | Example of reporting both absolute and relative effect sizes. (Adapted from table 3 of The OSIRIS Collaborative Group[242])

Primary outcome	Percentage (No)		Risk ratio (95% CI)	Risk difference (95% CI)
	Early administration (n=1344)	Delayed selective administration (n=1346)		
Death or oxygen dependence at "expected date of delivery"	31.9 (429)	38.2 (514)	0.84 (0.75 to 0.93)	−6.3 (−9.9 to −2.7)

Participants may sometimes not receive the full intervention, or some ineligible patients may have been randomly allocated in error. One widely recommended way to handle such issues is to analyse all participants according to their original group assignment, regardless of what subsequently occurred (see box 6). This "intention-to-treat" strategy is not always straightforward to implement. It is common for some patients not to complete a study—they may drop out or be withdrawn from active treatment—and thus are not assessed at the end. If the outcome is mortality, such patients may be included in the analysis based on register information, whereas imputation techniques may need to be used if other outcome data are missing. The term "intention-to-treat analysis" is often inappropriately used—for example, when those who did not receive the first dose of a trial drug are excluded from the analyses.[18]

Conversely, analysis can be restricted to only participants who fulfil the protocol in terms of eligibility, interventions, and outcome assessment. This analysis is known as an "on-treatment" or "per protocol" analysis. Excluding participants from the analysis can lead to erroneous conclusions. For example, in a trial that compared medical with surgical therapy for carotid stenosis, analysis limited to participants who were available for follow-up showed that surgery reduced the risk for transient ischaemic attack, stroke, and death. However, intention-to-treat analysis based on all participants as originally assigned did not show a superior effect of surgery.[214]

Intention-to-treat analysis is generally favoured because it avoids bias associated with non-random loss of participants.[215-217] Regardless of whether authors use the term "intention-to-treat," they should make clear which and how many participants are included in each analysis (see item 13). Non-compliance with assigned therapy may mean that the intention-to-treat analysis underestimates the potential benefit of the treatment, and additional analyses, such as a per protocol analysis, may therefore be considered.[218 219] It should be noted, however, that such analyses are often considerably flawed.[220]

In a review of 403 RCTs published in 10 leading medical journals in 2002, 249 (62%) reported the use of intention-to-treat analysis for their primary analysis. This proportion was higher for journals adhering to the CONSORT statement (70% *v* 48%). Among articles that reported the use of intention-to-treat analysis, only 39% actually analysed all participants as randomised, with more than 60% of articles having missing data in their primary analysis.[221] Other studies show similar findings.[18 222 223] Trials with no reported exclusions are methodologically weaker in other respects than those that report on some excluded participants,[173] strongly indicating that at least some researchers who have excluded participants do not report it. Another study found that reporting an intention-to-treat analysis was associated with other aspects of good study design and reporting, such as describing a sample size calculation.[224]

Item 17a. For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)

Examples—See tables 5 and 6.

Explanation—For each outcome, study results should be reported as a summary of the outcome in each group (for example, the number of participants with or without the event and the denominators, or the mean and standard deviation of measurements), together with the contrast between the groups, known as the effect size. For binary outcomes, the effect size could be the risk ratio (relative risk), odds ratio, or risk difference; for survival time data, it could be the hazard ratio or difference in median survival time; and for continuous data, it is usually the difference in means. Confidence intervals should be presented for the contrast between groups. A common error is the presentation of separate confidence intervals for the outcome in each group rather than for the treatment effect.[233] Trial results are often more clearly displayed in a table rather than in the text, as shown in tables 5 and 6.

For all outcomes, authors should provide a confidence interval to indicate the precision (uncertainty) of the estimate.[48 235] A 95% confidence interval is conventional, but occasionally other levels are used. Many journals require or strongly encourage the use of confidence intervals.[236] They are especially valuable in relation to differences that do not meet conventional statistical significance, for which they often indicate that the result does not rule out an important clinical difference. The use of confidence intervals has increased markedly in recent years, although not in all medical specialties.[233] Although P values may be provided in addition to confidence intervals, results should not be reported solely as P values.[237 238] Results should be reported for all planned primary and secondary end points, not just for analyses that were statistically significant or "interesting." Selective reporting within a study is a widespread and serious problem.[55 57] In trials in which interim analyses were performed, interpretation should focus on the final results at the close of the trial, not the interim results.[239]

For both binary and survival time data, expressing the results also as the number needed to treat for benefit or harm can be helpful (see item 21).[240 241]

Item 17b. For binary outcomes, presentation of both absolute and relative effect sizes is recommended

Example—"The risk of oxygen dependence or death was reduced by 16% (95% CI 25% to 7%). The absolute difference was −6.3% (95% CI −9.9% to −2.7%); early administration to an estimated 16 babies would therefore prevent 1 baby dying or being long-term dependent on oxygen" (also see table 7).[242]

Explanation—When the primary outcome is binary, both the relative effect (risk ratio (relative risk) or odds ratio) and the absolute effect (risk difference) should be reported (with confidence intervals), as neither the relative measure nor the absolute measure alone gives a complete picture of the effect and its implications. Different audiences may prefer either relative or absolute risk, but both doctors and lay people

tend to overestimate the effect when it is presented in terms of relative risk.[243-245] The size of the risk difference is less generalisable to other populations than the relative risk since it depends on the baseline risk in the unexposed group, which tends to vary across populations. For diseases where the outcome is common, a relative risk near unity might indicate clinically important differences in public health terms. In contrast, a large relative risk when the outcome is rare may not be so important for public health (although it may be important to an individual in a high risk category).

Item 18. Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory

Example—"On the basis of a study that suggested perioperative β-blocker efficacy might vary across baseline risk, we prespecified our primary subgroup analysis on the basis of the revised cardiac risk index scoring system. We also did prespecified secondary subgroup analyses based on sex, type of surgery, and use of an epidural or spinal anaesthetic. For all subgroup analyses, we used Cox proportional hazard models that incorporated tests for interactions, designated to be significant at p<0.05 ... Figure 3 shows the results of our prespecified subgroup analyses and indicates consistency of effects ... Our subgroup analyses were underpowered to detect the modest differences in subgroup effects that one might expect to detect if there was a true subgroup effect."[100]

Explanation—Multiple analyses of the same data create a risk for false positive findings.[246] Authors should resist the temptation to perform many subgroup analyses.[183 185 247] Analyses that were prespecified in the trial protocol (see item 24) are much more reliable than those suggested by the data, and therefore authors should report which analyses were prespecified. If subgroup analyses were undertaken, authors should report which subgroups were examined, why, if they were prespecified, and how many were prespecified. Selective reporting of subgroup analyses could lead to bias.[248] When evaluating a subgroup the question is not whether the subgroup shows a statistically significant result but whether the subgroup treatment effects are significantly different from each other. To determine this, a test of interaction is helpful, although the power for such tests is typically low. If formal evaluations of interaction are undertaken (see item 12b) they should be reported as the estimated difference in the intervention effect in each subgroup (with a confidence interval), not just as P values.

In one survey, 35 of 50 trial reports included subgroup analyses, of which only 42% used tests of interaction.[183] It was often difficult to determine whether subgroup analyses had been specified in the protocol. In another survey of surgical trials published in high impact journals, 27 of 72 trials reported 54 subgroup analyses, of which 91% were post hoc and only 6% of subgroup analyses used a test of interaction to assess whether a subgroup effect existed.[249]

Similar recommendations apply to analyses in which adjustment was made for baseline variables. If done, both unadjusted and adjusted analyses should be reported. Authors should indicate whether adjusted analyses, including the choice of variables to adjust for, were planned. Ideally, the trial protocol should state whether adjustment is made for nominated baseline variables by using analysis of covariance.[187] Adjustment for variables because they differ significantly at baseline is likely to bias the estimated treatment effect.[187] A survey found that unacknowledged discrepancies between protocols and publications were found for all 25 trials reporting subgroup analyses and for 23 of 28 trials reporting adjusted analyses.[92]

Item 19. All important harms or unintended effects in each group

For specific guidance see CONSORT for harms.[42]

Example—"The proportion of patients experiencing any adverse event was similar between the rBPI21 [recombinant bactericidal/permeability-increasing protein] and placebo groups: 168 (88.4%) of 190 and 180 (88.7%) of 203, respectively, and it was lower in patients treated with rBPI21 than in those treated with placebo for 11 of 12 body systems ... the proportion of patients experiencing a severe adverse event, as judged by the investigators, was numerically lower in the rBPI21 group than the placebo group: 53 (27.9%) of 190 versus 74 (36.5%) of 203 patients, respectively. There were only three serious adverse events reported as drug-related and they all occurred in the placebo group."[250]

Explanation—Readers need information about the harms as well as the benefits of interventions to make rational and balanced decisions. The existence and nature of adverse effects can have a major impact on whether a particular intervention will be deemed acceptable and useful. Not all reported adverse events observed during a trial are necessarily a consequence of the intervention; some may be a consequence of the condition being treated. Randomised trials offer the best approach for providing safety data as well as efficacy data, although they cannot detect rare harms.

Many reports of RCTs provide inadequate information on adverse events. A survey of 192 drug trials published from 1967 to 1999 showed that only 39% had adequate reporting of clinical adverse events and 29% had adequate reporting of laboratory defined toxicity.[72] More recently, a comparison between the adverse event data submitted to the trials database of the National Cancer Institute, which sponsored the trials, and the information reported in journal articles found that low grade adverse events were underreported in journal articles. High grade events (Common Toxicity Criteria grades 3 to 5) were reported inconsistently in the articles, and the information regarding attribution to investigational drugs was incomplete.[251] Moreover, a review of trials published in six general medical journals in 2006 to 2007 found that, although 89% of 133 reports mentioned adverse events, no information on severe adverse events and withdrawal of patients due to an adverse event was given on 27% and 48% of articles, respectively.[252]

An extension of the CONSORT statement has been developed to provide detailed recommendations on the reporting of harms in randomised trials.[42] Recommendations and examples of appropriate reporting are freely available from the CONSORT website (www.consort-statement.org). They complement the CONSORT 2010 Statement and should be consulted, particularly if the study of harms was a key objective. Briefly, if data on adverse events were collected, events should be listed and defined, with reference to standardised criteria where appropriate. The methods

used for data collection and attribution of events should be described. For each study arm the absolute risk of each adverse event, using appropriate metrics for recurrent events, and the number of participants withdrawn due to harms should be presented. Finally, authors should provide a balanced discussion of benefits and harms.[42]

Discussion

Item 20. Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses

Example—"The preponderance of male patients (85%) is a limitation of our study ... We used bare-metal stents, since drug-eluting stents were not available until late during accrual. Although the latter factor may be perceived as a limitation, published data indicate no benefit (either short-term or long-term) with respect to death and myocardial infarction in patients with stable coronary artery disease who receive drug-eluting stents, as compared with those who receive bare-metal stents."[253]

Explanation—The discussion sections of scientific reports are often filled with rhetoric supporting the authors' findings[254] and provide little measured argument of the pros and cons of the study and its results. Some journals have attempted to remedy this problem by encouraging more structure to authors' discussion of their results.[255 256] For example, *Annals of Internal Medicine* recommends that authors structure the discussion section by presenting (1) a brief synopsis of the key findings, (2) consideration of possible mechanisms and explanations, (3) comparison with relevant findings from other published studies (whenever possible including a systematic review combining the results of the current study with the results of all previous relevant studies), (4) limitations of the present study (and methods used to minimise and compensate for those limitations), and (5) a brief section that summarises the clinical and research implications of the work, as appropriate.[255] We recommend that authors follow these sensible suggestions, perhaps also using suitable subheadings in the discussion section.

Although discussion of limitations is frequently omitted from research reports,[257] identification and discussion of the weaknesses of a study have particular importance.[258] For example, a surgical group reported that laparoscopic cholecystectomy, a technically difficult procedure, had significantly lower rates of complications than the more traditional open cholecystectomy for management of acute cholecystitis.[259] However, the authors failed to discuss an obvious bias in their results. The study investigators had completed all the laparoscopic cholecystectomies, whereas 80% of the open cholecystectomies had been completed by trainees.

Authors should also discuss any imprecision of the results. Imprecision may arise in connection with several aspects of a study, including measurement of a primary outcome (see item 6a) or diagnosis (see item 4a). Perhaps the scale used was validated on an adult population but used in a paediatric one, or the assessor was not trained in how to administer the instrument.

The difference between statistical significance and clinical importance should always be borne in mind. Authors should particularly avoid the common error of interpreting a non-significant result as indicating equivalence of interventions. The confidence interval (see item 17a) provides valuable insight into whether the trial result is compatible with a clinically important effect, regardless of the P value.[120]

Authors should exercise special care when evaluating the results of trials with multiple comparisons. Such multiplicity arises from several interventions, outcome measures, time points, subgroup analyses, and other factors. In such circumstances, some statistically significant findings are likely to result from chance alone.

Item 21. Generalisability (external validity, applicability) of the trial findings

Examples—"As the intervention was implemented for both sexes, all ages, all types of sports, and at different levels of sports, the results indicate that the entire range of athletes, from young elite to intermediate and recreational senior athletes, would benefit from using the presented training programme for the prevention of recurrences of ankle sprain. By including non-medically treated and medically treated athletes, we covered a broad spectrum of injury severity. This suggests that the present training programme can be implemented in the treatment of all athletes. Furthermore, as it is reasonable to assume that ankle sprains not related to sports are comparable with those in sports, the programme could benefit the general population."[260]

"This replicates and extends the work of Clarke and colleagues and demonstrates that this CB (cognitive behavioural) prevention program can be reliably and effectively delivered in different settings by clinicians outside of the group who originally developed the intervention. The effect size was consistent with those of previously reported, single-site, indicated depression prevention studies and was robust across sites with respect to both depressive disorders and symptoms ... In this generalisability trial, we chose a comparison condition that is relevant to public health—usual care ... The sample also was predominantly working class to middle class with access to health insurance. Given evidence that CB therapy can be more efficacious for adolescents from homes with higher incomes, it will be important to test the effects of this prevention program with more economically and ethnically diverse samples."[261]

Explanation—External validity, also called generalisability or applicability, is the extent to which the results of a study can be generalised to other circumstances.[262] Internal validity, the extent to which the design and conduct of the trial eliminate the possibility of bias, is a prerequisite for external validity: the results of a flawed trial are invalid and the question of its external validity becomes irrelevant. There is no absolute external validity; the term is meaningful only with regard to clearly specified conditions that were not directly examined in the trial. Can results be generalised to an individual participant or groups that differ from those enrolled in the trial with regard to age, sex, severity of disease, and comorbid conditions? Are the results applicable to other drugs within a class of similar drugs, to a different dose, timing, and route of administration, and to different concomitant therapies? Can similar results be expected at the primary, secondary, and tertiary levels of care? What about the effect on related outcomes that were not assessed in the trial, and the importance of length of follow-up and duration of treatment, especially with respect to harms?[263]

External validity is a matter of judgment and depends on

the characteristics of the participants included in the trial, the trial setting, the treatment regimens tested, and the outcomes assessed.[5 136] It is therefore crucial that adequate information be described about eligibility criteria and the setting and location (see item 4b), the interventions and how they were administered (see item 5), the definition of outcomes (see item 6), and the period of recruitment and follow-up (see item 14). The proportion of control group participants in whom the outcome develops (control group risk) is also important. The proportion of eligible participants who refuse to enter the trial as indicated on the flowchart (see item 13) is relevant for the generalisability of the trial, as it may indicate preferences for or acceptability of an intervention. Similar considerations may apply to clinician preferences.[264 265]

Several issues are important when results of a trial are applied to an individual patient.[266-268] Although some variation in treatment response between an individual patient and the patients in a trial or systematic review is to be expected, the differences tend to be in magnitude rather than direction.

Although there are important exceptions,[268] therapies (especially drugs [269]) found to be beneficial in a narrow range of patients generally have broader application in actual practice. Frameworks for the evaluation of external validity have been proposed, including qualitative studies, such as in integral "process evaluations"[270] and checklists.[271] Measures that incorporate baseline risk when calculating therapeutic effects, such as the number needed to treat to obtain one additional favourable outcome and the number needed to treat to produce one adverse effect, are helpful in assessing the benefit-to-risk balance in an individual patient or group with characteristics that differ from the typical trial participant.[268 272 273] Finally, after deriving patient centred estimates for the potential benefit and harm from an intervention, the clinician must integrate them with the patient's values and preferences for therapy. Similar considerations apply when assessing the generalisability of results to different settings and interventions.

Item 22. Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

Example—"Studies published before 1990 suggested that prophylactic immunotherapy also reduced nosocomial infections in very-low-birth-weight infants. However, these studies enrolled small numbers of patients; employed varied designs, preparations, and doses; and included diverse study populations. In this large multicenter, randomised controlled trial, the repeated prophylactic administration of intravenous immune globulin failed to reduce the incidence of nosocomial infections significantly in premature infants weighing 501 to 1500 g at birth."[274]

Explanation—Readers will want to know how the present trial's results relate to those of other RCTs. This can best be achieved by including a formal systematic review in the results or discussion section of the report.[83 275-277] Such synthesis may be impractical for trial authors, but it is often possible to quote a systematic review of similar trials. A systematic review may help readers assess whether the results of the RCT are similar to those of other trials in the same topic area and whether participants are similar across studies. Reports of RCTs have often not dealt adequately with these points.[277] Bayesian methods can be used to statistically combine the trial data with previous evidence.[278]

We recommend that, at a minimum, the discussion should be as systematic as possible and be based on a comprehensive search, rather than being limited to studies that support the results of the current trial.[279]

Other information

Item 23. Registration number and name of trial registry

Example—"The trial is registered at ClinicalTrials.gov, number NCT00244842."[280]

Explanation—The consequences of non-publication of entire trials,[281 282] selective reporting of outcomes within trials, and of per protocol rather than intention-to-treat analysis have been well documented.[55 56 283] Covert redundant publication of clinical trials can also cause problems, particularly for authors of systematic reviews when results from the same trial are inadvertently included more than once.[284]

To minimise or avoid these problems there have been repeated calls over the past 25 years to register clinical trials at their inception, to assign unique trial identification numbers, and to record other basic information about the trial so that essential details are made publicly available.[285-288] Provoked by recent serious problems of withholding data,[289] there has been a renewed effort to register randomised trials. Indeed, the World Health Organisation states that "the registration of all interventional trials is a scientific, ethical and moral responsibility" (www.who.int/ictrp/en). By registering a randomised trial, authors typically report a minimal set of information and obtain a unique trial registration number.

In September 2004 the International Committee of Medical Journal Editors (ICMJE) changed their policy, saying that they would consider trials for publication only if they had been registered before the enrolment of the first participant.[290] This resulted in a dramatic increase in the number of trials being registered.[291] The ICMJE gives guidance on acceptable registries (www.icmje.org/faq.pdf).

In a recent survey of 165 high impact factor medical journals' instructions to authors, 44 journals specifically stated that all recent clinical trials must be registered as a requirement of submission to that journal.[292]

Authors should provide the name of the register and the trial's unique registration number. If authors had not registered their trial they should explicitly state this and give the reason.

Item 24. Where the full trial protocol can be accessed, if available

Example—"Full details of the trial protocol can be found in the Supplementary Appendix, available with the full text of this article at www.nejm.org."[293]

Explanation—A protocol for the complete trial (rather than a protocol of a specific procedure within a trial) is important because it pre-specifies the methods of the randomised trial, such as the primary outcome (see item 6a). Having a protocol can help to restrict the likelihood of undeclared post hoc changes to the trial methods and selective outcome reporting (see item 6b). Elements that may be important for inclusion in the protocol for a randomised trial are described elsewhere.[294]

There are several options for authors to consider ensur-

ing their trial protocol is accessible to interested readers. As described in the example above, journals reporting a trial's primary results can make the trial protocol available on their web site. Accessibility to the trial results and protocol is enhanced when the journal is open access. Some journals (such as *Trials*) publish trial protocols, and such a publication can be referenced when reporting the trial's principal results. Trial registration (see item 23) will also ensure that many trial protocol details are available, as the minimum trial characteristics included in an approved trial registration database includes several protocol items and results (www.who.int/ictrp/en). Trial investigators may also be able to post their trial protocol on a website through their employer. Whatever mechanism is used, we encourage all trial investigators to make their protocol easily accessible to interested readers.

Item 25. Sources of funding and other support (such as supply of drugs), role of funders

Examples—"Grant support was received for the intervention from Plan International and for the research from the Wellcome Trust and Joint United Nations Programme on HIV/AIDS (UNAIDS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."[295]

"This study was funded by GlaxoSmithKline Pharmaceuticals. GlaxoSmithKline was involved in the design and conduct of the study and provided logistical support during the trial. Employees of the sponsor worked with the investigators to prepare the statistical analysis plan, but the analyses were performed by the University of Utah. The manuscript was prepared by Dr Shaddy and the steering committee members. GlaxoSmithKline was permitted to review the manuscript and suggest changes, but the final decision on content was exclusively retained by the authors."[296]

Explanation—Authors should report the sources of funding for the trial, as this is important information for readers assessing a trial. Studies have showed that research sponsored by the pharmaceutical industry are more likely to produce results favouring the product made by the company sponsoring the research than studies funded by other sources.[297-300] A systematic review of 30 studies on funding found that research funded by the pharmaceutical industry had four times the odds of having outcomes favouring the sponsor than research funded by other sources (odds ratio 4.05, 95% confidence interval 2.98 to 5.51).[297] A large proportion of trial publications do not currently report sources of funding. The degree of underreporting is difficult to quantify. A survey of 370 drug trials found that 29% failed to report sources of funding.[301] In another survey, of PubMed indexed randomised trials published in December 2000, source of funding was reported for 66% of the 519 trials.[16]

The level of involvement by a funder and their influence on the design, conduct, analysis, and reporting of a trial varies. It is therefore important that authors describe in detail the role of the funders. If the funder had no such involvement, the authors should state so. Similarly, authors should report any other sources of support, such as supply and preparation of drugs or equipment, or in the analysis of data and writing of the manuscript.[302]

Reporting RCTs that did not have a two group parallel design

The primary focus of the CONSORT recommendations is RCTs with a parallel design and two treatment groups. Most RCTs have that design, but a substantial minority do not: 45% (233/519) of RCTs published in December 2000,[16] and 39% (242/616) in December 2006.[17]

Most of the CONSORT statement applies equally to all trial designs, but there are a few additional issues to address for each design. Before the publication of the revised CONSORT statement in 2001, the CONSORT Group decided to develop extensions to the main CONSORT statement relevant to specific trial designs. Extensions have been published relating to reporting of cluster randomised trials[40] and non-inferiority and equivalence trials.[39] Lack of resources has meant that other planned extensions have not been completed; they will cover trials with the following designs: multiarm parallel, factorial, crossover, within-person.

Authors reporting trials with a cluster design or using a non-inferiority or equivalence framework should consult the CONSORT recommendations in addition to those in this document. Here we make a few interim comments about the other designs. In each case, the trial design should be made clear in both the main text and the article's abstract.

Multiarm (>2 group) parallel group trials need the least modification of the standard CONSORT guidance. The flow diagram can be extended easily. The main differences from trials with two groups relate to clarification of how the study hypotheses relate to the multiple groups, and the consequent methods of data analysis and interpretation. For factorial trials, the possibility of interaction between the interventions generally needs to be considered. In addition to overall comparisons of participants who did or did not receive each intervention under study, investigators should consider also reporting results for each treatment combination.[303]

In crossover trials, each participant receives two (or more) treatments in a random order. The main additional issues to address relate to the paired nature of the data, which affect design and analysis.[304] Similar issues affect within-person comparisons, in which participants receive two treatments simultaneously (often to paired organs). Also, because of the risk of temporal or systemic carryover effects, respectively, in both cases the choice of design needs justification.

The CONSORT Group intends to publish extensions to CONSORT to cover all these designs. In addition, we will publish updates to existing guidance for cluster randomised trials and non-inferiority and equivalence trials to take account of this major update of the generic CONSORT guidance.

Discussion

Assessment of healthcare interventions can be misleading unless investigators ensure unbiased comparisons. Random allocation to study groups remains the only method that eliminates selection and confounding biases. Non-randomised trials tend to result in larger estimated treatment effects than randomised trials.[305 306]

Bias jeopardises even RCTs, however, if investigators carry out such trials improperly.[307] A recent systematic

review, aggregating the results of several methodological investigations, found that, for subjective outcomes, trials that used inadequate or unclear allocation concealment yielded 31% larger estimates of effect than those that used adequate concealment, and trials that were not blinded yielded 25% larger estimates.[153] As might be expected, there was a strong association between the two.

The design and implementation of an RCT require methodological as well as clinical expertise, meticulous effort,[143 308] and a high level of alertness for unanticipated difficulties. Reports of RCTs should be written with similarly close attention to reducing bias. Readers should not have to speculate; the methods used should be complete and transparent so that readers can readily differentiate trials with unbiased results from those with questionable results. Sound science encompasses adequate reporting, and the conduct of ethical trials rests on the footing of sound science.[309]

We hope this update of the CONSORT explanatory article will assist authors in using the 2010 version of CONSORT and explain in general terms the importance of adequately reporting of trials. The CONSORT statement can help researchers designing trials in future[310] and can guide peer reviewers and editors in their evaluation of manuscripts. Indeed, we encourage peer reviewers and editors to use the CONSORT checklist to assess whether authors have reported on these items. Such assessments will likely improve the clarity and transparency of published trials. Because CONSORT is an evolving document, it requires a dynamic process of continual assessment, refinement, and, if necessary, change, which is why we have this update of the checklist and explanatory article. As new evidence and critical comments accumulate, we will evaluate the need for future updates.

The first version of the CONSORT statement, from 1996, seems to have led to improvement in the quality of reporting of RCTs in the journals that have adopted it.[50-54]. Other groups are using the CONSORT template to improve the reporting of other research designs, such as diagnostic tests[311] and observational studies.[312]

The CONSORT website (www.consort-statement.org) has been established to provide educational material and a repository database of materials relevant to the reporting of RCTs. The site includes many examples from real trials, including all of the examples included in this article. We will continue to add good and bad examples of reporting to the database, and we invite readers to submit further suggestions by contacting us through the website. The CONSORT Group will continue to survey the literature to find relevant articles that address issues relevant to the reporting of RCTs, and we invite authors of any such articles to notify us about them. All of this information will be made accessible through the CONSORT website, which is updated regularly.

More than 400 leading general and specialty journals and biomedical editorial groups, including the ICMJE, World Association of Medical Journal Editors, and the Council of Science Editors, have given their official support to CONSORT. We invite other journals concerned about the quality of reporting of clinical trials to endorse the CONSORT statement and contact us through our website to let us know of their support. The ultimate benefactors of these collective efforts should be people who, for whatever reason, require intervention from the healthcare community.

We are grateful to Frank Davidoff and Tom Lang for their involvement in the 2001 version of CONSORT explanation and elaboration document. A special thanks to Mary Ocampo, the Ottawa CONSORT coordinator, who helped deliver this explanation and elaboration paper and the CONSORT statement.

The CONSORT Group contributors to CONSORT 2010: Douglas G Altman, Centre for Statistics in Medicine, University of Oxford, UK; Virginia Barbour, *PLoS Medicine*, UK; Jesse A Berlin, Johnson & Johnson Pharmaceutical Research and Development, USA; Isabelle Boutron, University Paris 7 Denis Diderot, Assistance Publique des Hôpitaux de Paris, INSERM, France; PJ Devereaux, McMaster University, Canada; Kay Dickersin, Johns Hopkins Bloomberg School of Public Health, USA; Diana Elbourne, London School of Hygiene & Tropical Medicine, UK; Susan Ellenberg, University of Pennsylvania School of Medicine, USA; Val Gebski, University of Sydney, Australia; Steven Goodman, *Journal of the Society for Clinical Trials*, USA; Peter C Gøtzsche, Nordic Cochrane Centre, Denmark; Trish Groves, *BMJ*, UK; Steven Grunberg, American Society of Clinical Oncology, USA; Brian Haynes, McMaster University, Canada; Sally Hopewell, Centre for Statistics in Medicine, University of Oxford, UK; Astrid James, *Lancet*; Peter Juhn, Johnson & Johnson, USA; Philippa Middleton, University of Adelaide, Australia; Don Minckler, University of California Irvine, USA; David Moher, Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Canada; Victor M Montori, Knowledge and Encounter Research Unit, Mayo Clinic College of Medicine, USA; Cynthia Mulrow, *Annals of Internal Medicine*, USA; Stuart Pocock, London School of Hygiene & Tropical Medicine, UK; Drummond Rennie, *JAMA*, USA; David L Schriger, *Annals of Emergency Medicine*, USA; Kenneth F Schulz, Family Health International, USA; Iveta Simera, EQUATOR Network, UK; Elizabeth Wager, Sideview, UK.

Funding: We gratefully acknowledge financial support from United Kingdom National Institute for Health Research; Canadian Institutes of Health Research; Presidents Fund, Canadian Institutes of Health Research; Johnson & Johnson; *BMJ*; and the American Society for Clinical Oncology.

1 Rennie D. CONSORT revised—improving the reporting of randomized trials. *JAMA* 2001;285:2006-7.
2 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
3 Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA* 1998;279:1489-91.
4 Kjaergard LL, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy. 7th Cochrane Colloquium, Rome, Italy 1999.
5 Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
6 Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, Peterson TC, Nyren O, Bradley LA, et al. Drug treatment of functional dyspepsia: a systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol* 1996;91:660-73.
7 Talley NJ, Owen BK, Boyce P, Paterson K. Psychological treatments for irritable bowel syndrome: a critique of controlled treatment trials. *Am J Gastroenterol* 1996;91:277-83.
8 Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000;136:381-5.
9 Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology* 1999;30:1134-8.
10 Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966;195:1123-8.
11 Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976. *BMJ* 1977;1:85-7.
12 Hall JC, Hill D, Watts JM. Misuse of statistical methods in the Australasian surgical literature. *Aust N Z J Surg* 1982;52:541-3.
13 Altman DG. Statistics in medical journals. *Stat Med* 1982;1:59-71.
14 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;317:426-32.
15 Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283-4.
16 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.
17 Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports

of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340:c723
18 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.
19 Lai TY, Wong VW, Lam RF, Cheng AC, Lam DS, Leung GM. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. *Ophthalmic Epidemiol* 2007;14:390-8.
20 Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le LJ, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-6.
21 Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA* 1998;280:247-9.
22 Altman DG. Randomisation. *BMJ* 1991;302:1481-2.
23 Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;272:125-8.
24 Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 1948;2:769-82.
25 Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol* 1998;41:245-56.
26 Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;1:421-9.
27 Armitage P. The role of randomization in clinical trials. *Stat Med* 1982;1:345-52.
28 Kleijnen J, Gøtzsche PC, Kunz R, Oxman AD, Chalmers I. So what's so special about randomisation. In: Maynard A, Chalmers I, eds. *Non-random reflections on health services research*. BMJ Books, 1997:93-106.
29 Chalmers I. Assembling comparison groups to assess the effects of health care. *J R Soc Med* 1997;90:379-86.
30 Nicolucci A, Grilli R, Alexanian AA, Apolone G, Torri V, Liberati A. Quality, evolution, and clinical implications of randomized, controlled trials on the treatment of lung cancer. A lost opportunity for meta-analysis. *JAMA* 1989;262:2101-7.
31 Ah-See KW, Molony NC. A qualitative assessment of randomized controlled trials in otolaryngology. *J Laryngol Otol* 1998;112:460-3.
32 Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.
33 Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;317:1181-4.
34 DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;306:1332-7.
35 A proposal for structured reporting of randomized controlled trials. The Standards of Reporting Trials Group. *JAMA* 1994;272:1926-31.
36 Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med* 1994;121:894-5.
37 Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers. *JAMA* 1995;273:1054-5.
38 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276:637-9.
39 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152-60.
40 Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328:702-8.
41 Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.
42 Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
43 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295-309.
44 Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 2006;144:364-7.
45 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.
46 Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996;276:1339-41.
47 Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;313:275-83.
48 Lang TA, Secic M. *How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers*. ACP, 1997.
49 Davidoff F. News from the International Committee of Medical Journal Editors. *Ann Intern Med* 2000;133:229-31.
50 Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263-7.
51 Egger M, Jüni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996-9.
52 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134:657-62.
53 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.
54 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
55 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
56 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by the Lancet. *Lancet* 2008;372:201.
57 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.
58 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
59 Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
60 Pocock SJ. *Clinical trials: a practical approach*. John Wiley, 1983.
61 Meinert CL. *Clinical trials: design, conduct and analysis*. Oxford University Press, 1986.
62 Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. 3rd ed. Springer, 1998.
63 Bolliger CT, Zellweger JP, Danielsson T, van Biljon X, Robidou A, Westin A, et al. Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety. *BMJ* 2000;321:329-33.
64 Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002;25:38-64.
65 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008;371:281-3.
66 The impact of open access upon public health. *PLoS Med* 2006;3:e252.
67 Harbourt AM, Knecht LS, Humphreys BL. Structured abstracts in MEDLINE, 1989-1991. *Bull Med Libr Assoc* 1995;83:190-5.
68 Harris AH, Standard S, Brunning JL, Casey SL, Goldberg JH, Oliver L, et al. The accuracy of abstracts in psychology journals. *J Psychol* 2002;136:141-8.
69 Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999;281:1110-1.
70 Ward LG, Kendrach MG, Price SO. Accuracy of abstracts for original research articles in pharmacy journals. *Ann Pharmacother* 2004;38:1173-7.
71 Gøtzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ* 2006;333:231-4.
72 Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
73 Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990;113:69-76.
74 Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ* 1994;150:1611-5.
75 Wager E, Middleton P. Technical editing of research reports in biomedical journals. *Cochrane Database Syst Rev* 2008;MR000002.
76 Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *J Inform Sci* 1996;22:349-56.
77 Gilligan D, Nicolson M, Smith I, Groen H, Dalesio O, Goldstraw P, et al. Preoperative chemotherapy in patients with resectable non-small cell lung cancer: results of the MRC LU22/NVALT 2/EORTC 08012 multicentre randomised trial and update of systematic review. *Lancet* 2007;369:1929-37.
78 Sandler AD, Sutton KA, DeWeese J, Girardi MA, Sheppard V, Bodfish JW. Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive developmental disorder. *N Engl J Med* 1999;341:1801-6.
79 World Medical Association. Declaration of Helsinki: ethical principle for medical research involving human subjects. 59th WMA General Assembly, Seoul 2008; www.wma.net/e/policy/b3.htm (accessed 2 June 2009).
80 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327:248-54.
81 Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials* 2005;2:218-29.
82 Savulescu J, Chalmers I, Blunt J. Are research ethics committees

behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996;313:1390-3.
83 Sadler LC, Davison T, McCowan LM. A randomised controlled trial and meta-analysis of active management of labour. *BJOG* 2000;107:909-15.
84 Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke* 1998;29:2203-10.
85 Blumer JL, Findling RL, Shih WJ, Soubrane C, Reed MD. Controlled clinical trial of zolpidem for the treatment of insomnia associated with attention-deficit/hyperactivity disorder in children 6 to 17 years of age. *Pediatrics* 2009;123:e770-e776.
86 Sabatine MS, Antman EM, Widimsky P, Ebrahim IO, Kiss RG, Saaiman A, et al. Otamixaban for the treatment of patients with non-ST-elevation acute coronary syndromes (SEPIA-ACS1 TIMI 42): a randomised, double-blind, active-controlled, phase 2 trial. *Lancet* 2009;374:787-95.
87 Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess* 2005;9:1-iv.
88 Gallo P, Krams M. PhRMA Working Group on adaptive designs, "White Paper." *Drug Information Journal* 2006;40:421-82.
89 Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, Muthen B, et al. Adaptive designs for randomized trials in public health. *Annu Rev Public Health* 2009;30:1-25.
90 Kelly PJ, Sooriyarachchi MR, Stallard N, Todd S. A practical comparison of group-sequential and adaptive designs. *J Biopharm Stat* 2005;15:719-38.
91 Pildal J, Chan AW, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005;330:1049.
92 Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
93 Ndekha MJ, van Oosterhout JJ, Zijlstra EE, Manary M, Saloojee H, Manary MJ. Supplementary feeding with either ready-to-use fortified spread or corn-soy blend in wasted adults starting antiretroviral therapy in Malawi: randomised, investigator blinded, controlled trial. *BMJ* 2009;338:1867-75.
94 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005;365:82-93.
95 Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. National Surgical Adjuvant Breast and Bowel Program. Pediatric Oncology Group. *J Clin Epidemiol* 1998;51:69-79.
96 Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol* 2000;53:973-9.
97 Gandhi M, Ameli N, Bacchetti P, Sharp GB, French AL, Young M, et al. Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS* 2005;19:1885-96.
98 Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery* 1996;119:466-72.
99 Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. *Arch Intern Med* 2008;168:133-5.
100 Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC, et al. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 2008;371:1839-47.
101 Rannou F, Dimet J, Boutron I, Baron G, Fayad F, Macé Y, et al. Splint for base-of-thumb osteoarthritis: a randomized trial. *Ann Intern Med* 2009;150:661-9.
102 Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;336:1472-4.
103 Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet* 2000;356:385-90.
104 McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 3rd ed. New York: Oxford University Press, 2006.
105 Streiner D, Norman C. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.
106 Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39.
107 Sanders C, Egger M, Donovan J, Tallon D, Frankel S. Reporting on quality of life in randomised controlled trials: bibliographic study. *BMJ* 1998;317:1191-4.
108 Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31-56.
109 Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249-52.
110 Jadad AR, Boyle M, Cunnigham C, Kim M, Schachar R. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report/Technology Assessment No. 11. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality. AHQR publication no. 00-E005; 1999.
111 Schachter HM, Pham B, King J, Langford S, Moher D. The efficacy and safety of methylphenidate in attention deficit disorder: A systematic review and meta-analyis. Prepared for the Therapeutics Initiative, Vancouver, B.C., and the British Columbia Ministry for Children and Families, 2000.
112 Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* 2001;357:1385-90.
113 Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
114 Vermeulen H, Hofland J, Legemate DA, Ubbink DT. Intravenous fluid restriction after major abdominal surgery: a randomized blinded clinical trial. *Trials* 2009;10:50.
115 Fox K, Ford I, Steg PG, Tendera M, Ferrari R. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *Lancet* 2008;372:807-16.
116 Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ* 1995;311:1145-8.
117 Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter. *PLoS Med* 2008;5:e4.
118 Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365:1348-53.
119 Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288:358-62.
120 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
121 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-4.
122 Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978;299:690-4.
123 Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.
124 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409-22.
125 Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121:200-6.
126 Galgiani JN, Catanzaro A, Cloud GA, Johnson RH, Williams PL, Mirels LF, et al. Comparison of oral fluconazole and itraconazole for progressive, nonmeningeal coccidioidomycosis. A randomized, double-blind trial. Mycoses Study Group. *Ann Intern Med* 2000;133:676-86.
127 Connolly SJ, Pogue J, Hart RG, Hohnloser SH, Pfeffer M, Chrolavicius S, et al. Effect of clopidogrel added to aspirin in patients with atrial fibrillation. *N Engl J Med* 2009;360:2066-78.
128 Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics* 1987;43:213-23.
129 Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med* 1985;4:521-6.
130 Pocock SJ. When to stop a clinical trial. *BMJ* 1992;305:235-40.
131 DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. *Lancet* 1999;354:1983-8.
132 Buyse M. Interim analyses, stopping rules and data monitoring in clinical trials in Europe. *Stat Med* 1993;12:509-20.
133 Sydes MR, Altman DG, Babiker AB, Parmar MK, Spiegelhalter DJ. Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study. *Clin Trials* 2004;1:48-59.
134 Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.
135 Coutinho IC, Ramos de Amorim MM, Katz L, Bandeira de Ferraz AA. Uterine exteriorization compared with in situ repair at cesarean delivery: a randomized controlled trial. *Obstet Gynecol* 2008;111:639-47.
136 Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. BMJ Books, 2001.
137 Creinin MD, Meyn LA, Borgatta L, Barnhart K, Jensen J, Burke AE, et al. Multicenter comparison of the contraceptive ring and patch: a randomized controlled trial. *Obstet Gynecol* 2008;111:267-77.
138 Tate DF, Jackvony EH, Wing RR. Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. *JAMA* 2003;289:1833-6.
139 Lachin JM. Properties of simple randomization in clinical trials. *Control Clin Trials* 1988;9:312-26.
140 Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;34:585-612.
141 Schulz KF, Grimes DA. *The Lancet handbook of essential concepts in clinical research*. Elsevier, 2006.
142 Altman DG, Bland JM. How to randomise. *BMJ* 1999;319:703-4.
143 Schulz KF. Subverting randomization in controlled trials. *JAMA*

1995;274:1456-8.
144 Enas GG, Enas NH, Spradlin CT, Wilson MG, Wiltse CG. Baseline comparability in clinical trials: prevention of poststudy anxiety. *Drug Information Journal* 1990;24:541-8.
145 Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* 1998;317:362-3.
146 Sinei SK, Schulz KF, Lamptey PR, Grimes DA, Mati JK, Rosenthal SM, et al. Preventing IUCD-related pelvic infection: the efficacy of prophylactic doxycycline at insertion. *Br J Obstet Gynaecol* 1990;97:412-9.
147 Radford JA, Landorf KB, Buchbinder R, Cook C. Effectiveness of low-Dye taping for the short-term treatment of plantar heel pain: a randomised trial. *BMC Musculoskelet Disord* 2006;7:64.
148 Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med* 1987;6:315-28.
149 Pocock SJ. Statistical aspects of clinical trial design. *Statistician* 1982;31:1-18.
150 Haag U. Technologies for automating randomized treatment assignment in clinical trials. *Drug Information Journal* 1998;32:11.
151 Piaggio G, Elbourne D, Schulz KF, Villar J, Pinol AP, Gülmezoglu AM. The reporting of methods for reducing and detecting bias: an example from the WHO Misoprostol Third Stage of Labour equivalence randomised controlled trial. *BMC Med Res Methodol* 2003;3:19.
152 Pildal J, Hróbjartsson A, Jórgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol* 2007;36:847-57.
153 Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.
154 McCandlish R, Bowler U, van Asten H, Berridge G, Winter C, Sames L, et al. A randomised controlled trial of care of the perineum during second stage of normal labour. *Br J Obstet Gynaecol* 1998;105:1262-72.
155 Webster J, Clarke S, Paterson D, Hutton A, van Dyk S, Gale C, et al. Routine care of peripheral intravenous catheters versus clinically indicated replacement: randomised controlled trial. *BMJ* 2008;337:a339.
156 Smith SA, Shah ND, Bryant SC, Christianson TJ, Bjornsen SS, Giesler PD, et al. Chronic care model and shared care in diabetes: randomized trial of an electronic decision support system. *Mayo Clin Proc* 2008;83:747-57.
157 Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med* 2009;360:859-73.
158 Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998;72:389-433.
159 Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL, et al. Effect of encouragement on walking test performance. *Thorax* 1984;39:818-22.
160 Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials* 1996;17:285-90.
161 Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. *JAMA* 1975;231:1038-42.
162 Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994;44:16-20.
163 Carley SD, Libetta C, Flavin B, Butler J, Tong N, Sammy I. An open prospective randomised trial to reduce the pain of blood glucose testing: ear versus thumb. *BMJ* 2000;321:20.
164 Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254-9.
165 Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ* 2000;321:504.
166 Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol* 2002;55:787-90.
167 Cheng K, Smyth RL, Motley J, O'Hea U, Ashby D. Randomized controlled trials in cystic fibrosis (1966-1997) categorized by time, design, and intervention. *Pediatr Pulmonol* 2000;29:1-7.
168 Lang T. Masking or blinding? An unscientific survey of mostly medical journal editors on the great debate. *Med Gen Med* 2000;2:E25.
169 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000-3.
170 Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006;3:360-5.
171 Meinert CL. Masked monitoring in clinical trials—blind stupidity? *N Engl J Med* 1998;338:1381-2.
172 Mills E, Prousky J, Raskin G, Gagnier J, Rachlis B, Montori VM, et al. The safety of over-the-counter niacin. A randomized placebo-controlled trial [ISRCTN18054903]. *BMC Clin Pharmacol* 2003;3:4.
173 Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742-4.
174 Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;328:432.
175 Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004;328:1136.
176 Astrup A, Rössner S, Van Gaal L, Rissanen A, Niskanen L, Al HM, et al. Effects of liraglutide in the treatment of obesity: a randomised, double-blind, placebo-controlled study. *Lancet* 2009;374:1606-16.
177 Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with confidence: confidence intervals and statistical guidelines*. 2nd ed. BMJ Books, 2000:171-90.
178 Altman DG, Bland JM. Statistics notes. Units of analysis. *BMJ* 1997;314:1874.
179 Bolton S. Independence and statistical inference in clinical trial designs: a tutorial review. *J Clin Pharmacol* 1998;38:408-12.
180 Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000;29:158-67.
181 Albert CM, Cook NR, Gaziano JM, Zaharris E, MacFadyen J, Danielson E, et al. Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: a randomized trial. *JAMA* 2008;299:2027-36.
182 Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ* 1996;313:862.
183 Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000;355:1064-9.
184 Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ* 1996;313:808.
185 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
186 Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am Heart J* 2000;139:745-51.
187 Altman DG. Adjustment for covariate imbalance. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. John Wiley, 1998:1000-5.
188 Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 2002;136:122-6.
189 Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201-10.
190 Bender R, Grouven U. Logistic regression models used in medical research are poorly presented. *BMJ* 1996;313:628.
191 Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol* 1999;93:1014-20.
192 Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;301:1410-2.
193 May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation* 1981;64:669-73.
194 Altman DG, Cuzick J, Peto J. More on zidovudine in asymptomatic HIV infection. *N Engl J Med* 1994;330:1758-9.
195 Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. Consolidated Standards of Reporting Trials. *JAMA* 1998;279:1487-9.
196 Grant AM, Wileman SM, Ramsay CR, Mowat NA, Krukowski ZH, Heading RC, et al. Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ* 2008;337:a2664.
197 van Loon AJ, Mantingh A, Serlier EK, Kroon G, Mooyaart EL, Huisjes HJ. Randomised controlled trial of magnetic-resonance pelvimetry in breech presentation at term. *Lancet* 1997;350:1799-804.
198 Brown MJ, Palmer CR, Castaigne A, de Leeuw PW, Mancia G, Rosenthal T, et al. Morbidity and mortality in patients randomised to double-blind treatment with a long-acting calcium-channel blocker or diuretic in the International Nifedipine GITS study: Intervention as a Goal in Hypertension Treatment (INSIGHT). *Lancet* 2000;356:366-72.
199 LaCroix AZ, Ott SM, Ichikawa L, Scholes D, Barlow WE. Low-dose hydrochlorothiazide and preservation of bone mineral density in older adults. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 2000;133:516-26.
200 Shuster JJ. Median follow-up in clinical trials. *J Clin Oncol* 1991;9:191-2.
201 Altman DG, de Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Br J Cancer* 1995;72:511-8.
202 Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med* 2005;2:e298.
203 Diggle L, Deeks J. Effect of needle length on incidence of local reactions to routine immunisation in infants aged 4 months: randomised controlled trial. *BMJ* 2000;321:931-3.
204 Pocock S, White I. Trials stopped early: too good to be true? *Lancet* 1999;353:943-4.
205 Hughes MD, Pocock SJ. Stopping rules and estimation problems in clinical trials. *Stat Med* 1988;7:1231-42.
206 Kiri A, Tonascia S, Meinert CL. Treatment effects monitoring committees and early stopping in large clinical trials. *Clin Trials* 2004;1:40-7.
207 Psaty BM, Rennie D. Stopping medical research to save money: a broken pact with researchers and patients. *JAMA* 2003;289:2128-31.
208 Temple R. FDA perspective on trials with interim efficacy evaluations. *Stat Med* 2006;25:3245-9.
209 Yusuf S, Teo K, Anderson C, Pogue J, Dyal L, Copland I, et al. Effects of

the angiotensin-receptor blocker telmisartan on cardiovascular events in high-risk patients intolerant to angiotensin-converting enzyme inhibitors: a randomised controlled trial. *Lancet* 2008;372:1174-83.
210 Senn S. Base logic: tests of baseline balance in randomized clinical trials. *Clin Res Regulatory Affairs* 1995;12:171-82.
211 Altman DG. Comparability of randomised groups. *Statistician* 1985;34:125-36.
212 Heit JA, Elliott CG, Trowbridge AA, Morrey BF, Gent M, Hirsh J. Ardeparin sodium for extended out-of-hospital prophylaxis against venous thromboembolism after total hip or knee replacement. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 2000;132:853-61.
213 Haderslev KV, Tjellesen L, Sorensen HA, Staun M. Alendronate increases lumbar spine bone mineral density in patients with Crohn's disease. *Gastroenterology* 2000;119:639-46.
214 Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA* 1970;211:1993-2003.
215 Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991;10:1595-605.
216 Lewis JA, Machin D. Intention to treat—who should use ITT? *Br J Cancer* 1993;68:647-50.
217 Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:526.
218 Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 1995;57:6-15.
219 Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med* 2000;19:1849-64.
220 Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
221 Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-6.
222 Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: who is in? Who is out? *J Fam Pract* 2002;51:969-71.
223 Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. *J Bone Joint Surg Am* 2009;91:2137-43.
224 Ruiz-Canela M, Martinez-González MA, de Irala-Estévez J. Intention to treat analysis is related to methodological quality. *BMJ* 2000;320:1007-8.
225 Altman DG. Missing outcomes in randomised trials: addressing the dilemma. *Open Med* 2009;3:e21-3.
226 Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004;1:368-76.
227 Streiner DL. Missing data and the trouble with LOCF. *Evid Based Ment Health* 2008;11:3-5.
228 Molnar FJ, Hutton B, Fergusson D. Does analysis using "last observation carried forward" introduce bias in dementia research? *CMAJ* 2008;179:751-3.
229 Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med* 2003;348:2136-7.
230 Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry* 2002;47:68-75.
231 Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharm Stat* 2008;7:93-106.
232 Abraha I, Montedori A, Romagnoli C. Modified intention to treat: frequency, definition and implication for clinical trials [abstract]. Sao Paulo, Brazil: XV Cochrane Colloquium, 2007: 86-7.
233 Altman DG. Confidence intervals in practice. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with confidence*. 2nd ed. BMJ Books, 2000:6-14.
234 van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, et al. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. *BMJ* 2009;339:b4074.
235 Altman DG. Clinical trials and meta-analyses. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with confidence*. 2nd ed. BMJ Books, 2000:120-38.
236 Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *Ann Intern Med* 1997;126:36-47.
237 Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* 1986;292:746-50.
238 Bailar JC, III, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med* 1988;108:266-73.
239 Bland JM. Quoting intermediate analyses can only mislead. *BMJ* 1997;314:1907-8.
240 Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452-4.
241 Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319:1492-5.
242 The OSIRIS Collaborative Group. Early versus delayed neonatal administration of a synthetic surfactant—the judgment of OSIRIS (open study of infants at high risk of or with respiratory insufficiency—the role of surfactant). *Lancet* 1992;340:1363-9.
243 Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexøe J, Nielsen JB. Laypersons' understanding of relative risk reductions: randomised cross-sectional study. *BMC Med Inform Decis Mak* 2008;8:31.
244 Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994;343:1209-11.
245 Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916-21.
246 Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679-84.
247 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93-8.
248 Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000;19:3325-36.
249 Bhandari M, Devereaux PJ, Li P, Mah D, Lim K, Schünemann HJ, et al. Misuse of baseline comparison tests and subgroup analyses in surgical trials. *Clin Orthop Relat Res* 2006;447:247-51.
250 Levin M, Quint PA, Goldstein B, Barton P, Bradley JS, Shemie SD, et al. Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: a randomised trial. rBPI21 Meningococcal Sepsis Study Group. *Lancet* 2000;356:961-7.
251 Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J Clin Oncol* 2006;24:3933-8.
252 Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-61.
253 Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med* 2007;356:1503-16.
254 Horton R. The rhetoric of research. *BMJ* 1995;310:985-7.
255 Annals of Internal Medicine. Information for authors. Available at www.annals.org (accessed 15 Jan 2008).
256 Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ* 1999;318:1224-5.
257 Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA* 1998;280:227-8.
258 Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol* 2007;60:324-9.
259 Kiviluoto T, Sirén J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet* 1998;351:321-5.
260 Hupperets MD, Verhagen EA, van Mechelen W. Effect of unsupervised home based proprioceptive training on recurrences of ankle sprain: randomised controlled trial. *BMJ* 2009;339:b2684.
261 Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TR, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA* 2009;301:2215-24.
262 Campbell D. Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957;54:297-312.
263 Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials* 2006;1:e9.
264 King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technol Assess* 2005;9:1-iv.
265 Djulbegovic B, Lacevic M, Cantor A, Fields KK, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635-8.
266 Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA* 1998;279:545-9.
267 Smith GD, Egger M. Who benefits from medical interventions? *BMJ* 1994;308:72-4.
268 McAlister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. BMJ Books, 2001.
269 Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess* 2005;9:iii-x, 1.
270 Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 2006;333:346-9.
271 Bornhöft G, Maxion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, et al. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. *BMC Med Res Methodol* 2006;6:56.
272 Laupacis A, Sackett DL, Roberts RS. An assessment of clinically

useful measures of the consequences of treatment. *N Engl J Med* 1988;318:1728-33.
273 Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;317:1309-12.
274 Fanaroff AA, Korones SB, Wright LL, Wright EC, Poland RL, Bauer CB, et al. A controlled trial of intravenous immune globulin to reduce nosocomial infections in very-low-birth-weight infants. National Institute of Child Health and Human Development Neonatal Research Network. *N Engl J Med* 1994;330:1107-13.
275 Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. First International Study of Infarct Survival Collaborative Group. *Lancet* 1986;2:57-66.
276 Gøtzsche PC, Gjørup I, Bonnén H, Brahe NE, Becker U, Burcharth F. Somatostatin v placebo in bleeding oesophageal varices: randomised trial and meta-analysis. *BMJ* 1995;310:1495-8.
277 Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. *J R Soc Med* 2007;100:187-90.
278 Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999;130:995-1004.
279 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
280 Papp K, Bissonnette R, Rosoph L, Wasel N, Lynde CW, Searles G, et al. Efficacy of ISA247 in plaque psoriasis: a randomised, multicentre, double-blind, placebo-controlled phase III study. *Lancet* 2008;371:1337-42.
281 Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;9:15-21.
282 Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;4:1-115.
283 Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547-61.
284 Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635-40.
285 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-41.
286 Chalmers I. From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med* 2006;99:337-41.
287 Tonks A. A clinical trials register for Europe. *BMJ* 2002;325:1314-5.
288 Dickersin K, Rennie D. Registering clinical trials. *JAMA* 2003;290:516-23.
289 Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet* 2004;363:1341-5.
290 De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Lancet* 2005;365:1827-9.
291 Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DA. Issues in the registration of clinical trials. *JAMA* 2007;297:2112-20.
292 Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'instructions to authors'. *Trials* 2008;9:20.
293 Russell JA, Walley KR, Singer J, Gordon AC, Hébert PC, Cooper DJ, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med* 2008;358:877-87.
294 Chan AW, Tetzlaff J, Altman D, Gøtzsche PC, Hróbjartsson A, Krleza-Jeric K, et al. The SPIRIT initiative: defining standard protocol items for randomised trials. Oral presentation at the 16th Cochrane Colloquium: Evidence in the era of globalisation; 2008 Oct 3-7; Freiburg, Germany [abstract]. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen* 2008;102:27.
295 Gregson S, Adamson S, Papaya S, Mundondo J, Nyamukapa CA, Mason PR, et al. Impact and process evaluation of integrated community and clinic-based HIV-1 control: a cluster-randomised trial in eastern Zimbabwe. *PLoS Med* 2007;4:e102.
296 Shaddy RE, Boucek MM, Hsu DT, Boucek RJ, Canter CE, Mahony L, et al. Carvedilol for children and adolescents with heart failure: a randomized controlled trial. *JAMA* 2007;298:1171-9.
297 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
298 Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *BMJ* 2002;325:249.
299 Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Med* 2007;4:e184.
300 Sismondo S. Pharmaceutical company funding and its consequences: a qualitative systematic review. *Contemp Clin Trials* 2008;29:109-13.
301 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
302 Ross JS, Hill KP, Egilman DS, Krumholz HM. Guest authorship and ghostwriting in publications related to rofecoxib: a case study of industry documents from rofecoxib litigation. *JAMA* 2008;299:1800-12.
303 McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003;289:2545-53.
304 Senn S. *Crossover trials in clinical research.* 2nd ed. Wiley, 2002.
305 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-173.
306 Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007;MR000012.
307 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001;357:373-80.
308 Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet* 1996;348:596-8.
309 Murray GD. Promoting good research practice. *Stat Methods Med Res* 2000;9:17-24.
310 Narahari SR, Ryan TJ, Aggithaya MG, Bose KS, Prasanna KS. Evidence-based approaches for the Ayurvedic traditional herbal formulations: toward an Ayurvedic CONSORT model. *J Altern Complement Med* 2008;14:769-76.
311 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003;138:40-4.
312 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573-7.
313 Tonino PA, De Bruyne B, Pijls NH, Siebert U, Ikeno F, van't Veer M, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N Engl J Med* 2009;360:213-24.

Towards minimum reporting standards for life scientists

The “minimum standards” working group[1]

Transparency in reporting benefits scientific communication on many levels. While specific needs and expectations vary across fields, the effective use of research findings relies on the availability of core information about research materials, data, and analysis. In December 2017, a working group of journal editors and experts in reproducibility convened to create the “minimum standards” working group. This working group aims to devise a set of minimum expectations that journals could ask their authors to meet, and will draw from the collective experience of journals implementing a range of different approaches designed to enhance reporting and reproducibility (e.g. STAR Methods), existing life science checklists (e.g. the Nature Research reporting summary), and the results of recent meta-research studying the efficacy of such interventions (e.g. Macleod et al. 2017; Han et al. 2017).

The working group will provide three key deliverables:

- A “minimum standards” framework setting out minimum expectations across four core areas of materials (including data and code), design, analysis and reporting (MDAR)
- A “minimum standards” checklist intended to operationalize the framework by serving as an implementation tool to aid authors in complying with journal policies, and editors and reviewers in assessing reporting and compliance with policies
- An “elaboration” document or user guide providing context for the “minimum standards” framework and checklist

While all three outputs are intended to provide tools to help journals, researchers and other stakeholders with adoption of the minimum standards framework, we do not intend to be prescriptive about the precise mechanism of implementation and we anticipate that in many cases they will be used as a yardstick within the context of an existing reporting system. Nevertheless, we hope these tools will provide a consolidated view to help raise reporting standards across the life sciences.

The working group will pilot these deliverables with a small number of journals and ask them to use the checklist within the editorial review and author revision process. Authors and editors will be asked to qualitatively evaluate the utility of the checklist. We will also seek out qualitative feedback on the framework, checklist, and elaboration document from a small stakeholder group of researchers and experts in reproducibility and transparent reporting. Feedback from this process will be used to consider next steps, which might include revisions to the content of the checklist. Following this process, journals will be free to use the checklist in ways that support their workflows.

The data and materials for this evaluation will be made available as the supplemental materials to this preprint after the initial pilot period, which is expected to end in September, 2019. Data will only be provided in aggregate form, that is, not parsed by individual article or journal, so as to respect the confidentiality of responses.

An advantage of aligning on minimum standards is consistency in policies and expectations across journals, which is beneficial for authors as they prepare papers for publication and for reviewers as they assess them. We also hope that other major stakeholders engaged in the research cycle, including institutional review bodies and funders, will see the value of agreeing on this type of reporting standard as a minimum expectation, as broad-based endorsement from an early stage in the research life cycle would provide important support for overall adoption and implementation.

---

[1] Karen Chambers (Wiley; kchamber@wiley.com), Andy Collings (eLife; a.collings@elifesciences.org), Chris Graf (Wiley; cgraf@wiley.com), Veronique Kiermer (Public Library of Science; vkiermer@plos.org; first corresponding author), David Mellor (Center for Open Science; david@cos.io), Malcolm Macleod (University of Edinburgh; Malcolm.macleod@ed.ac.uk), Sowmya Swaminathan (Nature Research/Springer Nature; s.swaminathan@us.nature.com, second corresponding author), Deborah Sweet (Cell Press/Elsevier; dsweet@cell.com), Valda Vinson (Science/AAAS; vvinson@aaas.org)

Guidelines and Guidance

Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

David Moher[1,2]*, Alessandro Liberati[3,4], Jennifer Tetzlaff[1], Douglas G. Altman[5], The PRISMA Group[¶]

1 Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, **2** Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada, **3** Università di Modena e Reggio Emilia, Modena, Italy, **4** Centro Cochrane Italiano, Istituto Ricerche Farmacologiche Mario Negri, Milan, Italy, **5** Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom

Introduction

Systematic reviews and meta-analyses have become increasingly important in health care. Clinicians read them to keep up to date with their field [1,2], and they are often used as a starting point for developing clinical practice guidelines. Granting agencies may require a systematic review to ensure there is justification for further research [3], and some health care journals are moving in this direction [4]. As with all research, the value of a systematic review depends on what was done, what was found, and the clarity of reporting. As with other publications, the reporting quality of systematic reviews varies, limiting readers' ability to assess the strengths and weaknesses of those reviews.

Several early studies evaluated the quality of review reports. In 1987, Mulrow examined 50 review articles published in four leading medical journals in 1985 and 1986 and found that none met all eight explicit scientific criteria, such as a quality assessment of included studies [5]. In 1987, Sacks and colleagues [6] evaluated the adequacy of reporting of 83 meta-analyses on 23 characteristics in six domains. Reporting was generally poor; between one and 14 characteristics were adequately reported (mean = 7.7; standard deviation = 2.7). A 1996 update of this study found little improvement [7].

In 1996, to address the suboptimal reporting of meta-analyses, an international group developed a guidance called the QUOROM Statement (*QU*ality *O*f *R*eporting *O*f *M*eta-analyses), which focused on the reporting of meta-analyses of randomized controlled trials [8]. In this article, we summarize a revision of these guidelines, renamed PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses), which have been updated to address several conceptual and practical advances in the science of systematic reviews (Box 1).

Terminology

The terminology used to describe a systematic review and meta-analysis has evolved over time. One reason for changing the name from QUOROM to PRISMA was the desire to encompass both systematic reviews and meta-analyses. We have adopted the definitions used by the Cochrane Collaboration [9]. A systematic review is a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies. Meta-analysis refers to the use of statistical techniques in a systematic review to integrate the results of included studies.

Developing the PRISMA Statement

A three-day meeting was held in Ottawa, Canada, in June 2005 with 29 participants, including review authors, methodologists, clinicians, medical editors, and a consumer. The objective of the Ottawa meeting was to revise and expand the QUOROM checklist and flow diagram, as needed.

The executive committee completed the following tasks, prior to the meeting: a systematic review of studies examining the quality of reporting of systematic reviews, and a comprehensive literature search to identify methodological and other articles that might inform the meeting, especially in relation to modifying checklist items. An international survey of review authors, consumers, and groups commissioning or using systematic reviews and meta-analyses was completed, including the International Network of Agencies for Health Technology Assessment (INAHTA) and the Guidelines International Network (GIN). The survey aimed to ascertain views of QUOROM, including the merits of the existing checklist items. The results of these activities were presented during the meeting and are summarized on the PRISMA Web site (http://www.prisma-statement.org/).

Only items deemed essential were retained or added to the checklist. Some additional items are nevertheless desirable, and review authors should include these, if relevant [10]. For example, it is useful to indicate whether the systematic review is an update [11] of a previous review, and to describe any changes in procedures from those described in the original protocol.

Citation: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed.1000097

Published July 21, 2009

Funding: PRISMA was funded by the Canadian Institutes of Health Research; Università di Modena e Reggio Emilia, Italy; Cancer Research UK; Clinical Evidence BMJ Knowledge; the Cochrane Collaboration; and GlaxoSmithKline, Canada. AL is funded, in part, through grants of the Italian Ministry of University (COFIN - PRIN 2002 prot. 2002061749 and COFIN - PRIN 2006 prot. 2006062298). DGA is funded by Cancer Research UK. DM is funded by a University of Ottawa Research Chair. None of the sponsors had any involvement in the planning, execution, or write-up of the PRISMA documents. Additionally, no funder played a role in drafting the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analyses; QUOROM, *QU*ality *O*f *R*eporting *O*f *M*eta-analyses.

* E-mail: dmoher@ohri.ca

¶ Membership of the PRISMA Group is provided in the Acknowledgments.

Provenance: Not commissioned; externally peer reviewed. In order to encourage dissemination of the PRISMA Statement, this article is freely accessible on the *PLoS Medicine* Web site (http://medicine.plosjournals.org/) and will be also published in the *Annals of Internal Medicine*, *BMJ*, *Journal of Clinical Epidemiology*, and *Open Medicine*. The authors jointly hold the copyright of this article. For details on further use, see the PRISMA Web site (http://www.prisma-statement.org/).

Box 1: Conceptual Issues in the Evolution from QUOROM to PRISMA

Completing a Systematic Review Is an Iterative Process The conduct of a systematic review depends heavily on the scope and quality of included studies: thus systematic reviewers may need to modify their original review protocol during its conduct. Any systematic review reporting guideline should recommend that such changes can be reported and explained without suggesting that they are inappropriate. The PRISMA Statement (Items 5, 11, 16, and 23) acknowledges this iterative process. Aside from Cochrane reviews, all of which should have a protocol, only about 10% of systematic reviewers report working from a protocol [22]. Without a protocol that is publicly accessible, it is difficult to judge between appropriate and inappropriate modifications.

Conduct and Reporting Research Are Distinct Concepts This distinction is, however, less straightforward for systematic reviews than for assessments of the reporting of an individual study, because the reporting and conduct of systematic reviews are, by nature, closely intertwined. For example, the failure of a systematic review to report the assessment of the risk of bias in included studies may be seen as a marker of poor conduct, given the importance of this activity in the systematic review process [37].

Study-Level Versus Outcome-Level Assessment of Risk of Bias For studies included in a systematic review, a thorough assessment of the risk of bias requires both a "study-level" assessment (e.g., adequacy of allocation concealment) and, for some features, a newer approach called "outcome-level" assessment. An outcome-level assessment involves evaluating the reliability and validity of the data for each important outcome by determining the methods used to assess them in each individual study [38]. The quality of evidence may differ across outcomes, even within a study, such as between a primary efficacy outcome, which is likely to be very carefully and systematically measured, and the assessment of serious harms [39], which may rely on spontaneous reports by investigators. This information should be reported to allow an explicit assessment of the extent to which an estimate of effect is correct [38].

Importance of Reporting Biases Different types of reporting biases may hamper the conduct and interpretation of systematic reviews. Selective reporting of complete studies (e.g., publication bias) [28] as well as the more recently empirically demonstrated "outcome reporting bias" within individual studies [40,41] should be considered by authors when conducting a systematic review and reporting its results. Though the implications of these biases on the conduct and reporting of systematic reviews themselves are unclear, some previous research has identified that selective outcome reporting may occur also in the context of systematic reviews [42].

Shortly after the meeting a draft of the PRISMA checklist was circulated to the group, including those invited to the meeting but unable to attend. A disposition file was created containing comments and revisions from each respondent, and the checklist was subsequently revised 11 times. The group approved the checklist, flow diagram, and this summary paper.

Although no direct evidence was found to support retaining or adding some items, evidence from other domains was believed to be relevant. For example, Item 5 asks authors to provide registration information about the systematic review, including a registration number, if available. Although systematic review registration is not yet widely available [12,13], the participating journals of the International Committee of Medical Journal Editors (ICMJE) [14] now require all clinical trials to be registered in an effort to increase transparency and accountability [15]. Those aspects are also likely to benefit systematic reviewers, possibly reducing the risk of an excessive number of reviews addressing the same question [16,17] and providing greater transparency when updating systematic reviews.

The PRISMA Statement

The PRISMA Statement consists of a 27-item checklist (Table 1; see also Text S1 for a downloadable Word template for researchers to re-use) and a four-phase flow diagram (Figure 1; see also Figure S1 for a downloadable Word template for researchers to re-use). The aim of the PRISMA Statement is to help authors improve the reporting of systematic reviews and meta-analyses. We have focused on randomized trials, but PRISMA can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions. PRISMA may also be useful for critical appraisal of published systematic reviews. However, the PRISMA checklist is not a quality assessment instrument to gauge the quality of a systematic review.

From QUOROM to PRISMA

The new PRISMA checklist differs in several respects from the QUOROM checklist, and the substantive specific changes are highlighted in Table 2. Generally, the PRISMA checklist "decouples" several items present in the QUOROM checklist and, where applicable, several checklist items are linked to improve consistency across the systematic review report.

The flow diagram has also been modified. Before including studies and providing reasons for excluding others, the review team must first search the literature. This search results in records. Once these records have been screened and eligibility criteria applied, a smaller number of articles will remain. The number of included articles might be smaller (or larger) than the number of studies, because articles may report on multiple studies and results from a particular study may be published in several articles. To capture this information, the PRISMA flow diagram now requests information on these phases of the review process.

Endorsement

The PRISMA Statement should replace the QUOROM Statement for those journals that have endorsed QUOROM. We hope that other journals will support PRISMA; they can do so by registering on the PRISMA Web site. To underscore to authors, and others, the importance of transparent reporting of systematic reviews, we encourage supporting journals to reference the PRISMA Statement and include the PRISMA Web address in their Instructions to Authors. We also invite editorial organizations to consider endorsing PRISMA and encourage authors to adhere to its principles.

The PRISMA Explanation and Elaboration Paper

In addition to the PRISMA Statement, a supporting Explanation and Elaboration document has been produced [18] following the style used for other reporting guidelines [19–21]. The process

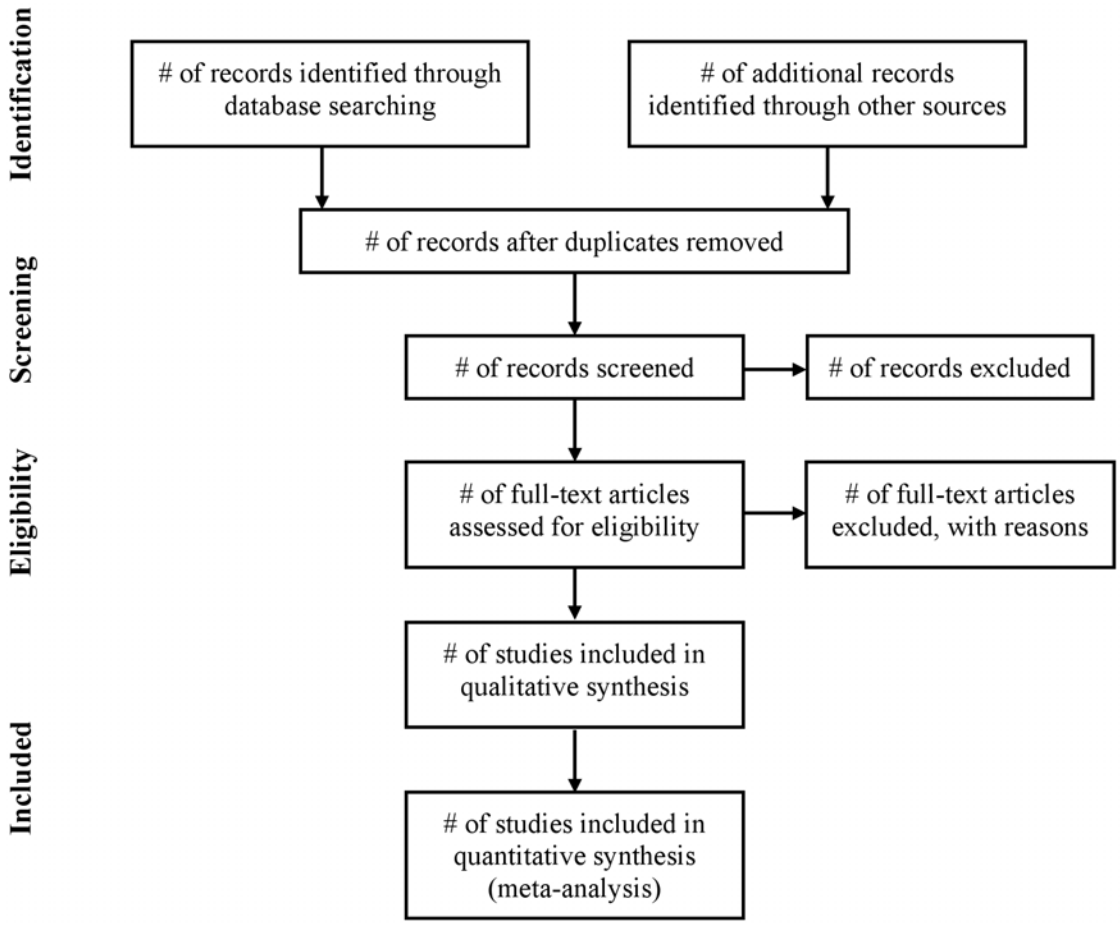


Figure 1. Flow of information through the different phases of a systematic review.
doi:10.1371/journal.pmed.1000097.g001

of completing this document included developing a large database of exemplars to highlight how best to report each checklist item, and identifying a comprehensive evidence base to support the inclusion of each checklist item. The Explanation and Elaboration document was completed after several face to face meetings and numerous iterations among several meeting participants, after which it was shared with the whole group for additional revisions and final approval. Finally, the group formed a dissemination subcommittee to help disseminate and implement PRISMA.

Discussion

The quality of reporting of systematic reviews is still not optimal [22–27]. In a recent review of 300 systematic reviews, few authors reported assessing possible publication bias [22], even though there is overwhelming evidence both for its existence [28] and its impact on the results of systematic reviews [29]. Even when the possibility of publication bias is assessed, there is no guarantee that systematic reviewers have assessed or interpreted it appropriately [30]. Although the absence of reporting such an assessment does not necessarily indicate that it was not done, reporting an assessment of possible publication bias is likely to be a marker of the thoroughness of the conduct of the systematic review.

Several approaches have been developed to conduct systematic reviews on a broader array of questions. For example, systematic reviews are now conducted to investigate cost-effectiveness [31], diagnostic [32] or prognostic questions [33], genetic associations [34], and policy making [35]. The general concepts and topics covered by PRISMA are all relevant to any systematic review, not just those whose objective is to summarize the benefits and harms of a health care intervention. However, some modifications of the checklist items or flow diagram will be necessary in particular circumstances. For example, assessing the risk of bias is a key concept, but the items used to assess this in a diagnostic review are likely to focus on issues such as the spectrum of patients and the verification of disease status, which differ from reviews of interventions. The flow diagram will also need adjustments when reporting individual patient data meta-analysis [36].

We have developed an explanatory document [18] to increase the usefulness of PRISMA. For each checklist item, this document contains an example of good reporting, a rationale for its inclusion, and supporting evidence, including references, whenever possible. We believe this document will also serve as a useful resource for those teaching systematic review methodology. We encourage journals to include reference to the explanatory document in their Instructions to Authors.

Like any evidence-based endeavor, PRISMA is a living document. To this end we invite readers to comment on the revised version, particularly the new checklist and flow diagram, through the PRISMA Web site. We will use such information to inform PRISMA's continued development.

Table 1. Checklist of items to include when reporting a systematic review or meta-analysis.

Section/Topic	#	Checklist Item	Reported on Page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see Item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

doi:10.1371/journal.pmed.1000097.t001

Table 2. Substantive specific changes between the QUOROM checklist and the PRISMA checklist (a tick indicates the presence of the topic in QUOROM or PRISMA).

Section/Topic	Item	QUOROM	PRISMA	Comment
Abstract		√	√	QUOROM and PRISMA ask authors to report an abstract. However, PRISMA is not specific about format.
Introduction	Objective		√	This new item (4) addresses the explicit question the review addresses using the PICO reporting system (which describes the participants, interventions, comparisons, and outcome(s) of the systematic review), together with the specification of the type of study design (PICOS); the item is linked to Items 6, 11, and 18 of the checklist.
Methods	Protocol		√	This new item (5) asks authors to report whether the review has a protocol and if so how it can be accessed.
Methods	Search	√	√	Although reporting the search is present in both QUOROM and PRISMA checklists, PRISMA asks authors to provide a full description of at least one electronic search strategy (Item 8). Without such information it is impossible to repeat the authors' search.
Methods	Assessment of risk of bias in included studies	√	√	Renamed from "quality assessment" in QUOROM. This item (12) is linked with reporting this information in the results (Item 19). The new concept of "outcome-level" assessment has been introduced.
Methods	Assessment of risk of bias across studies		√	This new item (15) asks authors to describe any assessments of risk of bias in the review, such as selective reporting within the included studies. This item is linked with reporting this information in the results (Item 22).
Discussion		√	√	Although both QUOROM and PRISMA checklists address the discussion section, PRISMA devotes three items (24–26) to the discussion. In PRISMA the main types of limitations are explicitly stated and their discussion required.
Funding			√	This new item (27) asks authors to provide information on any sources of funding for the systematic review.

doi:10.1371/journal.pmed.1000097.t002

Supporting Information

Figure S1 Flow of information through the different phases of a systematic review (downloadable template document for researchers to re-use).
Found at: doi:10.1371/journal.pmed.1000097.s001 (0.08 MB DOC)

Text S1 Checklist of items to include when reporting a systematic review or meta-analysis (downloadable template document for researchers to re-use).
Found at: doi:10.1371/journal.pmed.1000097.s002 (0.04 MB DOC)

Acknowledgments

The following people contributed to the PRISMA Statement: Doug Altman, DSc, Centre for Statistics in Medicine (Oxford, UK); Gerd Antes, PhD, University Hospital Freiburg (Freiburg, Germany); David Atkins, MD, MPH, Health Services Research and Development Service, Veterans Health Administration (Washington, D. C., US); Virginia Barbour, MRCP, DPhil, *PLoS Medicine* (Cambridge, UK); Nick Barrowman, PhD, Children's Hospital of Eastern Ontario (Ottawa, Canada); Jesse A. Berlin, ScD, Johnson & Johnson Pharmaceutical Research and Development (Titusville, New Jersey, US); Jocalyn Clark, PhD, *PLoS Medicine* (at the time of writing, *BMJ*, London, UK); Mike Clarke, PhD, UK Cochrane Centre (Oxford, UK) and School of Nursing and Midwifery, Trinity College (Dublin, Ireland); Deborah Cook, MD, Departments of Medicine, Clinical Epidemiology and Biostatistics, McMaster University (Hamilton, Canada); Roberto D'Amico, PhD, Università di Modena e Reggio Emilia (Modena, Italy) and Centro Cochrane Italiano, Istituto Ricerche Farmacologiche Mario Negri (Milan, Italy); Jonathan J. Deeks, PhD, University of Birmingham (Birmingham, UK); P. J. Devereaux, MD, PhD, Departments of Medicine, Clinical Epidemiology and Biostatistics, McMaster University (Hamilton, Canada); Kay Dickersin, PhD, Johns Hopkins Bloomberg School of Public Health (Baltimore, Maryland, US); Matthias Egger, MD, Department of Social and Preventive Medicine, University of Bern (Bern, Switzerland); Edzard Ernst, MD, PhD, FRCP, FRCP(Edin), Peninsula Medical School (Exeter, UK); Peter C. Gøtzsche, MD, MSc, The Nordic Cochrane Centre (Copenhagen, Denmark); Jeremy Grimshaw, MBChB, PhD, FRCFP, Ottawa Hospital Research Institute (Ottawa, Canada); Gordon Guyatt, MD, Departments of Medicine, Clinical Epidemiology and Biostatistics, McMaster University (Hamilton, Canada); Julian Higgins, PhD, MRC Biostatistics Unit (Cambridge, UK); John P. A. Ioannidis, MD, University of Ioannina Campus (Ioannina, Greece); Jos Kleijnen, MD, PhD, Kleijnen Systematic Reviews Ltd (York, UK) and School for Public Health and Primary Care (CAPHRI), University of Maastricht (Maastricht, Netherlands); Tom Lang, MA, Tom Lang Communications and Training (Davis, California, US); Alessandro Liberati, MD, Università di Modena e Reggio Emilia (Modena, Italy) and Centro Cochrane Italiano, Istituto Ricerche Farmacologiche Mario Negri (Milan, Italy); Nicola Magrini, MD, NHS Centre for the Evaluation of the Effectiveness of Health Care – CeVEAS (Modena, Italy); David McNamee, PhD, *The Lancet* (London, UK); Lorenzo Moja, MD, MSc, Centro Cochrane Italiano, Istituto Ricerche Farmacologiche Mario Negri (Milan, Italy); David Moher, PhD, Ottawa Methods Centre, Ottawa Hospital Research Institute (Ottawa, Canada); Cynthia Mulrow, MD, MSc, *Annals of Internal Medicine* (Philadelphia, Pennsylvania, US); Maryann Napoli, Center for Medical Consumers (New York, New York, US); Andy Oxman, MD, Norwegian Health Services Research Centre (Oslo, Norway); Ba' Pham, MMath, Toronto Health Economics and Technology Assessment Collaborative (Toronto, Canada) (at the time of the first meeting of the group, GlaxoSmithKline Canada, Mississauga, Canada); Drummond Rennie, MD, FRCP, FACP, University of California San Francisco (San Francisco, California, US); Margaret Sampson, MLIS, Children's Hospital of Eastern Ontario (Ottawa, Canada); Kenneth F. Schulz, PhD, MBA, Family Health International (Durham, North Carolina, US); Paul G. Shekelle, MD, PhD, Southern California Evidence Based Practice Center (Santa Monica, California, US); Jennifer Tetzlaff, BSc, Ottawa Methods Centre, Ottawa Hospital Research Institute (Ottawa, Canada); David Tovey, FRCGP, The Cochrane Library, Cochrane Collaboration

(Oxford, UK) (at the time of the first meeting of the group, *BMJ*, London, UK); Peter Tugwell, MD, MSc, FRCPC, Institute of Population Health, University of Ottawa (Ottawa, Canada).

Author Contributions

ICMJE criteria for authorship read and met: DM AL JT DGA. Wrote the first draft of the paper: DM AL DGA. Contributed to the writing of the paper: DM AL JT DGA. Participated in regular conference calls, identified the participants, secured funds, planned the meeting, participated in the meeting, and drafted the manuscript: DM AL DGA. Participated in identifying the evidence base for PRISMA, refining the checklist, and drafting the manuscript: JT. Agree with the recommendations: DM AL JT DGA.

References

1. Oxman AD, Cook DJ, Guyatt GH (1994) Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 272: 1367–1371.
2. Swingler GH, Volmink J, Ioannidis JP (2003) Number of published systematic reviews and global burden of disease: Database analysis. BMJ 327: 1083–1084.
3. Canadian Institutes of Health Research (2006) Randomized controlled trials registration/application checklist (12/2006). Available: http://www.cihr-irsc.gc.ca/e/documents/rct_reg_e.pdf. Accessed 19 May 2009.
4. Young C, Horton R (2005) Putting clinical trials into context. Lancet 366: 107.
5. Mulrow CD (1987) The medical review article: State of the science. Ann Intern Med 106: 485–488.
6. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC (1987) Meta-analysis of randomized controlled trials. New Engl J Med 316: 450–455.
7. Sacks HS, Reitman D, Pagano D, Kupelnick B (1996) Meta-analysis: An update. Mt Sinai J Med 63: 216–224.
8. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, et al. (1994) Improving the quality of reporting of meta-analysis of randomized controlled trials: The QUOROM statement. Lancet 354: 1896–1900.
9. Green S, Higgins J, eds (2005) Glossary. Cochrane handbook for systematic reviews of interventions 4.2.5. The Cochrane Collaboration. Available: http://www.cochrane.org/resources/glossary.htm. Accessed 19 May 2009.
10. Strech D, Tilburt J (2008) Value judgments in the analysis and synthesis of evidence. J Clin Epidemiol 61: 521–524.
11. Moher D, Tsertsvadze A (2006) Systematic reviews: When is an update an update? Lancet 367: 881–883.
12. University of York (2009) Centre for Reviews and Dissemination. Available: http://www.york.ac.uk/inst/crd/. Accessed 19 May 2009.
13. The Joanna Briggs Institute (2008) Protocols & work in progress. Available: http://www.joannabriggs.edu.au/pubs/systematic_reviews_prot.php. Accessed 19 May 2009.
14. De Angelis C, Drazan JM, Frizelle FA, Haug C, Hoey J, et al. (2004) Clinical trial registration: A statement from the International Committee of Medical Journal Editors. CMAJ 171: 606–607.
15. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, et al. (2004) Selective serotonin reuptake inhibitors in childhood depression: Systematic review of published versus unpublished data. Lancet 363: 1341–1345.
16. Bagshaw SM, McAlister FA, Manns BJ, Ghali WA (2006) Acetylcysteine in the prevention of contrast-induced nephropathy: A case study of the pitfalls in the evolution of evidence. Arch Intern Med 166: 161–166.
17. Biondi-Zoccai GG, Lotrionte M, Abbate A, Testa L, Remigi E, et al. (2006) Compliance with QUOROM and quality of reporting of overlapping meta-analyses on the role of acetylcysteine in the prevention of contrast associated nephropathy: Case study. BMJ 332: 202–209.
18. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche P, et al. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Med 6: e1000100. doi:10.1371/journal.pmed.1000100.
19. Altman DG, Schulz KR, Moher D, Egger M, Davidoff F, et al. (2001) The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. Ann Intern Med 134: 663–694.
20. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD explanation and elaboration. Ann Intern Med 138: W1–W12.
21. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, et al. (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. Ann Intern Med 147: W163–W194.
22. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG (2007) Epidemiology and reporting characteristics of systematic reviews. PLoS Med 4: e78. doi:10.1371/journal.pmed.0040078.
23. Bhandari M, Morrow F, Kulkarni AV, Tornetta P (2001) Meta-analyses in orthopaedic surgery: A systematic review of their methodologies. J Bone Joint Surg Am 83-A: 15–24.
24. Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH (2001) Evaluating the quality of systematic reviews in the emergency medicine literature. Ann Emerg Med 38: 518–526.
25. Richards D (2004) The quality of systematic reviews in dentistry. Evid Based Dent 5: 17.
26. Choi PT, Halpern SH, Malik N, Jadad AR, Tramer MR, et al. (2001) Examining the evidence in anesthesia literature: A critical appraisal of systematic reviews. Anesth Analg 92: 700–709.
27. Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB (2005) A systematic evaluation of the quality of meta-analyses in the critical care literature. Crit Care 9: R575–R582.
28. Dickersin K (2005) Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ, Borenstein M, eds. Publication bias in meta-analysis-Prevention, assessment and adjustments. Chichester (UK): John Wiley & Sons. pp 11–33.
29. Sutton AJ (2005) Evidence concerning the consequences of publication and related biases. In: Rothstein HR, Sutton AJ, Borenstein M, eds. Publication bias in meta-analysis-Prevention, assessment and adjustments. Chichester (UK): John Wiley & Sons. pp 175–192.
30. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I (2006) The case of the misleading funnel plot. BMJ 333: 597–600.
31. Ladabaum U, Chopra CL, Huang G, Scheiman JM, Chernew ME, et al. (2001) Aspirin as an adjunct to screening for prevention of sporadic colorectal cancer: A cost-effectiveness analysis. Ann Intern Med 135: 769–781.
32. Deeks JJ (2001) Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ 323: 157–162.
33. Altman DG (2001) Systematic reviews of evaluations of prognostic variables. BMJ 323: 224–228.
34. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29: 306–309.
35. Lavis J, Davies H, Oxman A, Denis J, Golden-Biddle K, et al. (2005) Towards systematic reviews that inform health care management and policy-making. J Health Serv Res Policy 10: 35–48.
36. Stewart LA, Clarke MJ (1995) Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. Stat Med 14: 2057–2079.
37. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, et al. (2005) Assessment of methodological quality of primary studies by systematic reviews: Results of the metaquality cross sectional study. BMJ 330: 1053–1055.
38. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, et al. (2008) GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336: 924–926.
39. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, et al. (2006) An official ATS statement: Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. Am J Respir Crit Care Med 174: 605–614.
40. Chan AW, Hrobjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. JAMA 291: 2457–2465.
41. Chan AW, Krleza-Jeric K, Schmid I, Altman DG (2004) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ 171: 735–740.
42. Silagy CA, Middleton P, Hopewell S (2002) Publishing protocols of systematic reviews: Comparing what was done to what was planned. JAMA 287: 2831–2834.

SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials

An-Wen Chan, MD, DPhil; Jennifer M. Tetzlaff, MSc; Douglas G. Altman, DSc; Andreas Laupacis, MD; Peter C. Gøtzsche, MD, DrMedSci; Karmela Krleža-Jerić, MD, DSc; Asbjørn Hróbjartsson, PhD; Howard Mann, MD; Kay Dickersin, PhD; Jesse A. Berlin, ScD; Caroline J. Doré, BSc; Wendy R. Parulekar, MD; William S.M. Summerskill, MBBS; Trish Groves, MBBS; Kenneth F. Schulz, PhD; Harold C. Sox, MD; Frank W. Rockhold, PhD; Drummond Rennie, MD; and David Moher, PhD

The protocol of a clinical trial serves as the foundation for study planning, conduct, reporting, and appraisal. However, trial protocols and existing protocol guidelines vary greatly in content and quality. This article describes the systematic development and scope of SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013, a guideline for the minimum content of a clinical trial protocol.

The 33-item SPIRIT checklist applies to protocols for all clinical trials and focuses on content rather than format. The checklist recommends a full description of what is planned; it does not prescribe how to design or conduct a trial. By providing guidance for key content, the SPIRIT recommendations aim to facilitate the drafting of high-quality protocols. Adherence to SPIRIT would also enhance the transparency and completeness of trial protocols for the benefit of investigators, trial participants, patients, sponsors, funders, research ethics committees or institutional review boards, peer reviewers, journals, trial registries, policymakers, regulators, and other key stakeholders.

Ann Intern Med. 2013;158:200-207. www.annals.org
For author affiliations, see end of text.
This article was published at www.annals.org on 8 January 2013.

The protocol of a clinical trial plays a key role in study planning, conduct, interpretation, oversight, and external review by detailing the plans from ethics approval to dissemination of results. A well-written protocol facilitates an appropriate assessment of scientific, ethical, and safety issues before a trial begins; consistency and rigor of trial conduct; and full appraisal of the conduct and results after trial completion. The importance of protocols has been emphasized by journal editors (1–6), peer reviewers (7–10), researchers (11–15), and public advocates (16).

Despite the central role of protocols, a systematic review revealed that existing guidelines for protocol content vary greatly in their scope and recommendations, seldom describe how the guidelines were developed, and rarely cite broad stakeholder involvement or empirical evidence to support their recommendations (17). These limitations may partly explain why an opportunity exists to improve the quality of protocols. Many protocols for randomized trials do not adequately describe the primary outcomes (inadequate for 25% of trials) (18, 19), treatment allocation methods (inadequate for 54% to 79%) (20, 21), use of blinding (inadequate for 9% to 34%) (21, 22), methods for reporting adverse events (inadequate for 41%) (23), components of sample size calculations (inadequate for 4% to 40%) (21, 24), data analysis plans (inadequate for 20% to 77%) (21, 24–26), publication policies (inadequate for 7%) (27), and roles of sponsors and investigators in study design or data access (inadequate for 89% to 100%) (28, 29). The problems that underlie these protocol deficiencies may in turn lead to avoidable protocol amendments, poor trial conduct, and inadequate reporting in trial publications (15, 30).

In response to these gaps in protocol content and guidance, we launched the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) initiative in 2007. This international project aims to improve the completeness of trial protocols by producing evidence-based recommendations for a minimum set of items to be addressed in protocols. The SPIRIT 2013 Statement includes a 33-item checklist (**Table 1**) and diagram (**Figure**). An associated explanatory paper (SPIRIT 2013 Explanation and Elaboration) (31) details the rationale and supporting evidence for each checklist item, along with guidance and model examples from actual protocols.

Development of the SPIRIT 2013 Statement

The SPIRIT 2013 Statement was developed in broad consultation with 115 key stakeholders, including trial investigators ($n = 30$); health care professionals ($n = 31$); methodologists ($n = 34$); statisticians ($n = 16$); trial coordinators ($n = 14$); journal editors ($n = 15$); and representatives from the research ethics community ($n = 17$), industry and nonindustry funders ($n = 7$), and regulatory agencies ($n = 3$), whose roles are not mutually exclusive. As detailed later, the SPIRIT guideline was developed through 2 systematic reviews, a formal Delphi consensus process, 2 face-to-face consensus meetings, and pilot-testing (32).

The SPIRIT checklist evolved through several iterations. The process began with a preliminary checklist of 59 items derived from a systematic review of existing protocol guidelines (17). In 2007, 96 expert panelists from 17 low- ($n = 1$), middle- ($n = 6$), and high-income ($n = 10$) countries refined this initial checklist over 3 iterative Delphi consensus survey rounds by e-mail (33). Panelists rated each item on a scale of 1 (not important) to 10 (very important), suggested new items, and provided comments that were circulated in subsequent rounds. Items with a median score of 8 or higher in the final round were included, whereas those rated 5 or lower were excluded.

Table 1. **SPIRIT 2013 Checklist: Recommended Items to Address in a Clinical Trial Protocol and Related Documents***

Section/Item	Item Number	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry.
	2b	All items from the World Health Organization Trial Registration Data Set (**Appendix Table**, available at www.annals.org)
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material, and other support
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors
	5b	Name and contact information for the trial sponsor
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities
	5d	Composition, roles, and responsibilities of the coordinating center, steering committee, end point adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see item 21a for DMC)
Introduction		
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention
	6b	Explanation for choice of comparators
Objectives	7	Specific objectives or hypotheses
Trial design	8	Description of trial design, including type of trial (e.g., parallel group, crossover, factorial, single group), allocation ratio, and framework (e.g., superiority, equivalence, noninferiority, exploratory)
Methods		
Participants, interventions, and outcomes		
Study setting	9	Description of study settings (e.g., community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centers and individuals who will perform the interventions (e.g., surgeons, psychotherapists)
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (e.g., drug dose change in response to harms, participant request, or improving/worsening disease)
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (e.g., drug tablet return, laboratory tests)
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (e.g., systolic blood pressure), analysis metric (e.g., change from baseline, final value, time to event), method of aggregation (e.g., median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended
Participant timeline	13	Time schedule of enrollment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (**Figure**).
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations
Recruitment	15	Strategies for achieving adequate participant enrollment to reach target sample size
Assignment of interventions (for controlled trials)		
Allocation		
Sequence generation	16a	Method of generating the allocation sequence (e.g., computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (e.g., blocking) should be provided in a separate document that is unavailable to those who enroll participants or assign interventions.
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (e.g., central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned
Implementation	16c	Who will generate the allocation sequence, who will enroll participants, and who will assign participants to interventions
Blinding (masking)	17a	Who will be blinded after assignment to interventions (e.g., trial participants, care providers, outcome assessors, data analysts), and how
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial

Continued on following page

Table 1—**Continued**

Section/Item	Item Number	Description
Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (e.g., duplicate measurements, training of assessors) and a description of study instruments (e.g., questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol.
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (e.g., double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol.
Statistical methods	20a	Statistical methods for analyzing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol.
	20b	Methods for any additional analyses (e.g., subgroup and adjusted analyses)
	20c	Definition of analysis population relating to protocol nonadherence (e.g., as-randomized analysis), and any statistical methods to handle missing data (e.g., multiple imputation)
Monitoring		
Data monitoring	21a	Composition of DMC; summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed.
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor
Ethics and dissemination		
Research ethics approval	24	Plans for seeking REC/IRB approval
Protocol amendments	25	Plans for communicating important protocol modifications (e.g., changes to eligibility criteria, outcomes, analyses) to relevant parties (e.g., investigators, RECs/IRBs, trial participants, trial registries, journals, regulators)
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorized surrogates, and how (see item 32)
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site
Access to data	29	Statement of who will have access to the final trial data set, and disclosure of contractual agreements that limit such access for investigators
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, health care professionals, the public, and other relevant groups (e.g., via publication, reporting in results databases, or other data-sharing arrangements), including any publication restrictions
	31b	Authorship eligibility guidelines and any intended use of professional writers
	31c	Plans, if any, for granting public access to the full protocol, participant-level data set, and statistical code
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorized surrogates
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable

DMC = data monitoring committee; IRB = institutional review board; REC = research ethics committee; SPIRIT = Standard Protocol Items: Recommendations for Interventional Trials.

* It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation and Elaboration (31) for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group and is reproduced with permission.

Items rated between 5 and 8 were retained for further discussion at the consensus meetings.

After the Delphi survey, 16 members of the SPIRIT Group (named as authors of this paper) met in December 2007 in Ottawa, Ontario, Canada, and 14 members met in September 2009 in Toronto, Ontario, Canada, to review the survey results, discuss controversial items, and refine the draft checklist. After each meeting, the revised checklist was recirculated to the SPIRIT Group for additional feedback.

A second systematic review identified empirical evidence about the relevance of specific protocol items to trial conduct or risk of bias. The results of this review informed the decision to include or exclude items on the SPIRIT

checklist. This review also provided the evidence base of studies cited in the SPIRIT 2013 Explanation and Elaboration paper (31). Some items had little or no identified empirical evidence (for example, the title) and are included in the checklist on the basis of a strong pragmatic or ethical rationale.

Finally, we pilot-tested the draft checklist in 2010 and 2011 with University of Toronto graduate students who used the document to develop trial protocols as part of a master's-level course on clinical trial methods. Their feedback on the content, format, and usefulness of the checklist was obtained through an anonymous survey and incorporated into the final SPIRIT checklist.

Definition of a Clinical Trial Protocol

Although every study requires a protocol, the precise definition of a protocol varies among individual investigators, sponsors, and other stakeholders. For the SPIRIT initiative, the protocol is defined as a document that provides sufficient detail to enable understanding of the background, rationale, objectives, study population, interventions, methods, statistical analyses, ethical considerations, dissemination plans, and administration of the trial; replication of key aspects of trial methods and conduct; and appraisal of the trial's scientific and ethical rigor from ethics approval to dissemination of results.

The protocol is more than a list of items. It should be a cohesive document that provides appropriate context and narrative to fully understand the elements of the trial. For example, the description of a complex intervention may need to include training materials and figures to enable replication by persons with appropriate expertise.

The full protocol must be submitted for approval by an institutional review board (IRB) or research ethics committee (34). It is recommended that trial investigators or sponsors address the SPIRIT checklist items in the protocol before submission. If the details for certain items have not yet been finalized, then this should be stated in the protocol and the items updated as they evolve.

The protocol is a "living" document that is often modified during the trial. A transparent audit trail with dates of important changes in trial design and conduct is an essential part of the scientific record. Trial investigators and sponsors are expected to adhere to the protocol as approved by the IRB and to document amendments made in the most recent protocol version. Important protocol amendments should be reported to IRBs and trial registries as they occur and subsequently be described in trial reports.

Scope of the SPIRIT 2013 Statement

The SPIRIT 2013 Statement applies to the content of a clinical trial protocol, including its appendices. A clinical trial is a prospective study in which 1 or more interventions are assigned to human participants to assess the effects on health-related outcomes. The primary scope of SPIRIT 2013 relates to randomized trials, but the same considerations substantially apply to all types of clinical trials, regardless of study design, intervention, or topic.

The SPIRIT 2013 Statement provides guidance for minimum protocol content. Certain circumstances may warrant additional protocol items. For example, a factorial study design may require specific justification; crossover trials have unique statistical considerations, such as carryover effects; and industry-sponsored trials may have additional regulatory requirements.

Figure. **Example template of recommended content for the schedule of enrollment, interventions, and assessments.**

	Study Period							
	Enrollment	Allocation	Postallocation					Closeout
Time point*	$-t_1$	0	t_1	t_2	t_3	t_4	etc.	t_x
Enrollment:								
Eligibility screen	X							
Informed consent	X							
[List other procedures]	X							
Allocation		X						
Interventions:								
[Intervention A]			◆	—	◆			
[Intervention B]			X		X			
[List other study groups]			◆	—	—	◆		
Assessments:								
[List baseline variables]	X	X						
[List outcome variables]				X		X	etc.	X
[List other data variables]			X	X	X	X	etc.	X

Recommended content can be displayed using various schematic formats. See SPIRIT 2013 Explanation and Elaboration (31) for examples. This template is copyrighted by the SPIRIT Group and is reproduced with permission. SPIRIT = Standard Protocol Items: Recommendations for Interventional Trials.
* List specific time points in this row.

The protocol and its appendices are often the sole repository of detailed information relevant to every SPIRIT checklist item. Using existing trial protocols, we have been able to identify model examples of every item (31), which illustrates the feasibility of addressing all checklist items in a single protocol document. For some trials, relevant details may appear in related documents, such as statistical analysis plans, case record forms, operations manuals, or investigator contracts (35, 36). In these instances, the protocol should outline the key principles and refer to the separate documents so that their existence is known.

The SPIRIT 2013 Statement primarily relates to the content of the protocol rather than its format, which is often subject to local regulations, traditions, or standard operating procedures. Nevertheless, adherence to certain formatting conventions, such as a table of contents; section headings; glossary; list of abbreviations; list of references; and a schematic schedule of enrollment, interventions, and assessments, will facilitate protocol review (**Figure**).

Finally, the intent of SPIRIT 2013 is to promote transparency and a full description of what is planned—not to prescribe how a trial should be designed or conducted. The checklist should not be used to judge trial quality, because the protocol of a poorly designed trial may address all checklist items by fully describing its inadequate design features. Nevertheless, the use of SPIRIT 2013 may improve the validity and success of trials by reminding investigators about important issues to consider during the planning stages.

Relation to Existing Clinical Trial Guidance

With its systematic development process, consultation with international stakeholders, and explanatory paper citing relevant empirical evidence (31), SPIRIT 2013 builds on other international guidance applicable to clinical trial protocols. It adheres to the ethical principles mandated by the 2008 Declaration of Helsinki, particularly the requirement that the protocol address specific ethical considerations, such as competing interests (34).

In addition, SPIRIT 2013 encompasses the protocol items recommended by the International Conference on Harmonisation Good Clinical Practice E6 guidance, written in 1996 for clinical trials whose data are intended for submission to regulatory authorities (37). The SPIRIT Statement builds on the Good Clinical Practice guidance by providing additional recommendations on specific key protocol items (for example, allocation concealment, trial registration, and consent processes). In contrast to SPIRIT, the Good Clinical Practice guidance used informal consensus methods, has unclear contributorship, and lacks citation of supporting empirical evidence (38).

The SPIRIT 2013 Statement also supports trial registration requirements from the World Health Organization (39), the International Committee of Medical Journal Editors (40), legislation pertaining to ClinicalTrials.gov (41), the European Commission (42), and others. For example, item 2b of the SPIRIT checklist recommends that the protocol list the World Health Organization Trial Registration Data Set (**Appendix Table**, available at www.annals.org), which is the minimum amount of information that the International Committee of Medical Journal Editors mandates for trial registries. Having this data set in its own protocol section is intended not only to serve as a form of trial summary but also to help improve the quality of information in registry entries. Registration-specific data could be easily identified in the protocol section and copied into the registry fields. In addition, protocol amendments applicable to this section could prompt investigators to update their registry data.

The SPIRIT 2013 Statement mirrors applicable items from CONSORT 2010 (Consolidated Standards of Reporting Trials) (43). Consistent wording and structure used for items common to both checklists will facilitate the transition from a SPIRIT-based protocol to a final report based on CONSORT. The SPIRIT Group has also engaged leaders of other initiatives relevant to protocol standards, such as trial registries, the Clinical Data Interchange Standards Consortium Protocol Representation Group, and Pragmatic Randomized Controlled Trials in HealthCare, to align international efforts in promoting transparency and high-quality protocol content.

Potential Effect

An extensive range of stakeholders could benefit from widespread use of the SPIRIT 2013 Statement and its explanatory paper (**Table 2**). Pilot-testing and informal feedback have shown that it is particularly valuable for trial investigators when they draft their protocols. It can also serve as an informational resource for new investigators, peer reviewers, and IRB members.

There is also potential benefit for trial implementation. The excessive delay from the time of protocol development to ethics approval and the start of participant recruitment remains a major concern for clinical trials (44). Improved completeness of protocols could help increase the efficiency of protocol review by reducing avoidable queries to investigators about incomplete or unclear information. With full documentation of key information and increased awareness of important considerations before the trial begins, the use of SPIRIT may also help to reduce the number and burden of subsequent protocol amendments—many of which can be avoided with careful protocol drafting and development (15). Widespread adoption of SPIRIT 2013 as a single standard by IRBs, funding agencies, regulatory agencies, and journals could simplify the work of trial investigators and sponsors, who could fulfill the common application requirements of multiple stakeholders with a single SPIRIT-based protocol. Better protocols would also help trial personnel to implement the study as the protocol authors intended.

Table 2. **Potential Benefits and Proposed Stakeholder Actions for Supporting Adherence to SPIRIT 2013**

Stakeholder	Proposed Actions	Potential Benefits
Clinical trial groups, investigators, sponsors	Adopt SPIRIT as standard guidance Use as tool for writing protocols	Improved quality, completeness, and consistency of protocol content Enhanced understanding of rationale and issues to consider for key protocol items Increased efficiency of protocol review
Research ethics committees/institutional review boards, funding agencies, regulatory agencies	Mandate or encourage adherence to SPIRIT for submitted protocols Use as training tool	Improved quality, completeness, and consistency of protocol submissions Increased efficiency of review and reduction in queries about protocol requirements
Educators	Use SPIRIT checklist and explanatory paper as a training tool	Enhanced understanding of the rationale and issues to consider for key protocol items
Patients, trial participants, policymakers	Advocate use of SPIRIT by trial investigators and sponsors	Improved protocol content relevant to transparency, accountability, critical appraisal, and oversight
Trial registries	Encourage SPIRIT-based protocols Register full protocols to accompany results disclosure	Improved quality of registry records Prompt for trialists to update registry record when SPIRIT checklist item 2b (Registration Data Set) is updated Improved quality, completeness, and consistency of protocol content for registries that house full protocols and results
Journal editors and publishers	Endorse SPIRIT as standard guidance for published and unpublished protocols Include reference to SPIRIT in instructions for authors Ask that protocols be submitted with manuscripts, circulate them to peer reviewers, and encourage authors to make them available as Web appendices	Improved quality, completeness, and consistency of protocol content Enhanced peer review of trial manuscripts through improved protocol content, which can be used to assess protocol adherence and selective reporting Improved transparency and interpretation of trials by readers

SPIRIT = Standard Protocol Items: Recommendations for Interventional Trials.

Furthermore, adherence to SPIRIT 2013 could help ensure that protocols contain the requisite information for critical appraisal and trial interpretation. High-quality protocols can provide important information about trial methods and conduct that is not available from journals or trial registries (45–47). As a transparent record of the researchers' original intent, comparisons of protocols with final trial reports can help to identify selective reporting of results and undisclosed amendments (48), such as changes to primary outcomes (19, 49). However, clinical trial protocols are not generally accessible to the public (45). The SPIRIT 2013 Statement will have a greater effect when protocols are publicly available to facilitate full evaluation of trial validity and applicability (11, 12, 14, 50).

The SPIRIT 2013 guideline needs the support of key stakeholders to achieve its greatest impact (**Table 2**), as seen with widely adopted reporting guidelines, such as CONSORT (51). We will post the names of organizations that have endorsed SPIRIT 2013 on the SPIRIT Web site (www.spirit-statement.org) and provide resources to facilitate implementation. Widespread adoption of the SPIRIT recommendations can help improve protocol drafting, content, and implementation; facilitate registration, efficiency, and appraisal of trials; and ultimately enhance transparency for the benefit of patient care.

From Women's College Research Institute, Women's College Hospital, and Keenan Research Centre at the Li Ka Shing Knowledge Institute of St. Michael's Hospital, University of Toronto, Toronto, Ontario, Canada; Ottawa Methods Centre, Ottawa Hospital Research Institute, Ethics Office, Canadian Institutes of Health Research, and University of Ottawa, Ottawa, Ontario, Canada; Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom; Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark; University of Utah School of Medicine, Salt Lake City, Utah; Center for Clinical Trials, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; Janssen Research & Development, Janssen Pharmaceutical Companies of Johnson & Johnson, Titusville, New Jersey; UK Medical Research Council Clinical Trials Unit, *The Lancet*, and *BMJ*, London, United Kingdom; NCIC Clinical Trials Group, Cancer Research Institute, Queen's University, Kingston, Ontario, Canada; Quantitative Sciences, FHI 360, and GlaxoSmithKline, Research Triangle Park, North Carolina; The Dartmouth Institute for Health Policy and Clinical Practice, The Geisel School of Medicine at Dartmouth, Hanover, New Hampshire; The PR Lee Institute for Health Policy Studies, University of California, San Francisco, San Francisco, California.

Disclaimer: Dr. Krleža-Jerić was formerly employed by the Canadian Institutes of Health Research (Knowledge Translation Branch), and Dr. Parulekar is affiliated with the NCIC Clinical Trials Group. The funders otherwise had no input into the design and conduct of the project; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. Dr. Berlin is employed by the Janssen Pharmaceutical Companies of Johnson & Johnson, Dr. Sox is Editor Emeritus of *Annals of Internal Medicine*, and Dr. Rockhold is employed by GlaxoSmithKline.

Acknowledgment: The authors thank Drs. Mona Loufty and Patricia Parkin for pilot-testing the SPIRIT checklist with their graduate course students. The authors also acknowledge the participation of Dr. Geneviève Dubois-Flynn in the 2009 SPIRIT meeting.

Grant Support: Financial support for the SPIRIT meetings was provided by the Canadian Institutes of Health Research (grant DET-106068), National Cancer Institute of Canada (now Canadian Cancer Society Research Institute), and Canadian Agency for Drugs and Technologies in Health. The Canadian Institutes of Health Research has also funded ongoing dissemination activities (grant MET-117434).

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M12-1905.

Requests for Single Reprints: An-Wen Chan, MD, DPhil, Women's College Research Institute, Women's College Hospital, University of Toronto, 790 Bay Street, Toronto, Ontario M5G 1N8, Canada; e-mail, anwen.chan@utoronto.ca.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Rennie D.** Trial registration: a great idea switches from ignored to irresistible. JAMA. 2004;292:1359-62. [PMID: 15355937]
2. Strengthening the credibility of clinical research [Editorial]. Lancet. 2010;375:1225. [PMID: 20382309]
3. **Summerskill W, Collingridge D, Frankish H.** Protocols, probity, and publication. Lancet. 2009;373:992. [PMID: 19304003]
4. **Jones G, Abbasi K.** Trial protocols at the BMJ [Editorial]. BMJ. 2004;329:1360. [PMID: 15591544]
5. **Groves T.** Let SPIRIT take you . . . towards much clearer trial protocols. BMJ Group Blogs. 25 September 2009. Accessed at http://blogs.bmj.com/bmj/2009/09/25/trish-groves-let-spirit-take-you-towards-much-clearer-trial-protocols/ on 1 October 2012.
6. **Altman DG, Furberg CD, Grimshaw JM, Rothwell PM.** Lead editorial: trials-using the opportunities of electronic publishing to improve the reporting of randomised trials [Editorial]. Trials. 2006;7:6. [PMID: 16556322]
7. **Turner EH.** A taxpayer-funded clinical trials registry and results database [Editorial]. PLoS Med. 2004;1:e60. [PMID: 15562322]
8. **Coultas D.** Ethical considerations in the interpretation and communication of clinical trial results. Proc Am Thorac Soc. 2007;4:194-8. [PMID: 17494731]
9. **Siegel JP.** Editorial review of protocols for clinical trials [Letter]. N Engl J Med. 1990;323:1355. [PMID: 2215630]
10. **Murray GD.** Research governance must focus on research training. BMJ. 2001;322:1461-2.
11. **Chan A-W.** Access to clinical trial data [Editorial]. BMJ. 2011;342:d80. [PMID: 21228020]
12. **Miller JD.** Registering clinical trial results: the next step. JAMA. 2010;303:773-4. [PMID: 20179288]
13. **Krleža-Jerić K, Chan A-W, Dickersin K, Sim I, Grimshaw J, Gluud C.** Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). BMJ. 2005;330:956-8. [PMID: 15845980]
14. **Lassere M, Johnson K.** The power of the protocol. Lancet. 2002;360:1620-2. [PMID: 12457782]
15. **Getz KA, Zuckerman R, Cropp AB, Hindle AL, Krauss R, Kaitin KI.** Measuring the incidence, causes, and repercussions of protocol amendments. Drug Inf J. 2011;45:265-75.
16. Public Citizen Health Research Group v. Food and Drug Administration, 964 F Supp. 413 (DDC 1997).
17. **Tetzlaff JM, Chan A-W, Kitchen J, Sampson M, Tricco A, Moher D.** Guidelines for randomized clinical trial protocol content: a systematic review. Syst Rev. 2012;1:43. [PMID: 23006870]
18. **Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG.** Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA. 2004;291:2457-65. [PMID: 15161896]
19. **Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR.** Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. BMJ. 2011;342:c7153. [PMID: 21212122]
20. **Pildal J, Chan A-W, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC.** Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ. 2005;330:1049. [PMID: 15817527]
21. **Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A.** Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. J Clin Epidemiol. 2012;65:602-9. [PMID: 22424985]
22. **Hróbjartsson A, Pildal J, Chan A-W, Haahr MT, Altman DG, Gøtzsche PC.** Reporting on blinding in trial protocols and corresponding publications was often inadequate but rarely contradictory. J Clin Epidemiol. 2009;62:967-73. [PMID: 19635403]
23. **Scharf O, Colevas AD.** Adverse event reporting in publications compared with sponsor database for cancer clinical trials. J Clin Oncol. 2006;24:3933-8. [PMID: 16921045]
24. **Chan A-W, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG.** Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ. 2008;337:a2299. [PMID: 19056791]
25. **Al-Marzouki S, Roberts I, Evans S, Marshall T.** Selective reporting in clinical trials: analysis of trial protocols accepted by *The Lancet* [Letter]. Lancet. 2008;372:201. [PMID: 18640445]
26. **Hernández AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI.** Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. Neurosurgery. 2005;57:1244-53. [PMID: 16331173]
27. **Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan A-W.** Constraints on publication rights in industry-initiated clinical trials [Letter]. JAMA. 2006;295:1645-6. [PMID: 16609085]
28. **Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan A-W.** Ghost authorship in industry-initiated randomised trials. PLoS Med. 2007;4:e19. [PMID: 17227134]
29. **Lundh A, Krogsbøll LT, Gøtzsche PC.** Access to data in industry-sponsored trials [Letter]. Lancet. 2011;378:1995-6. [PMID: 22153200]
30. **Hopewell S, Dutton S, Yu LM, Chan A-W, Altman DG.** The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010;340:c723. [PMID: 20332510]
31. **Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al.** SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586.
32. **Moher D, Schulz KF, Simera I, Altman DG.** Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7:e1000217. [PMID: 20169112]
33. **Tetzlaff JM, Moher D, Chan A-W.** Developing a guideline for clinical trial protocol content: Delphi consensus survey. Trials. 2012;13:176. [PMID: 23006145]
34. **World Medical Association.** WMA Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects. Accessed at www.wma.net/en/30publications/10policies/b3/index.html on 1 October 2012.
35. **International Conference on Harmonisation.** ICH Harmonised Tripartite Guideline: General Considerations for Clinical Trials: E8. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 17 July 1997. Accessed at www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E8/Step4/E8_Guideline.pdf on 1 October 2012.
36. **International Conference on Harmonisation.** ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials: E9. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 5 February 1998. Accessed at www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf on 1 October 2012.
37. **International Conference on Harmonisation.** ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice: E6. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 10 June 1996. Accessed at www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf on 1 October 2012.
38. **Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG.** The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet. 2005;366:172-4. [PMID: 16005342]
39. **Sim I, Chan A-W, Gülmezoglu AM, Evans T, Pang T.** Clinical trial registration: transparency is the watchword. Lancet. 2006;367:1631-3. [PMID: 16714166]
40. **Laine C, De Angelis C, Delamothe T, Drazen JM, Frizelle FA, Haug C, et al.** Clinical trial registration: looking back and moving ahead [Editorial]. Ann Intern Med. 2007;147:275-7. [PMID: 17548404]
41. Food and Drug Administration Amendments Act of 2007, HR 2580, 110th Congress, 1st Sess, Title VIII, §801 (2007). Expanded Clinical Trial Registry

Data Bank. Accessed at www.govtrack.us/congress/billtext.xpd?bill=h110-3580 on 1 October 2012.
42. **European Commission.** Communication from the Commission regarding the guideline on the data fields contained in the clinical trials database provided for in Article 11 of Directive 2001/20/EC to be included in the database on medicinal products provided for in Article 57 of Regulation (EC) No 726/2004 (2008/C 168/02). Official Journal of the European Union. 2008;51:3-4.
43. **Schulz KF, Altman DG, Moher D; CONSORT Group.** CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med. 2010;152:726-32. [PMID: 20335313]
44. **National Research Council.** A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program. Washington, DC: National Academies Pr; 2010.
45. **Chan A-W.** Out of sight but not out of mind: how to search for unpublished clinical trial evidence. BMJ. 2012;344:d8013. [PMID: 22214892]
46. **Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T.** Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. BMJ. 2012;344:d8141. [PMID: 22214759]
47. **Reveiz L, Chan A-W, Krleža-Jerić K, Granados CE, Pinart M, Etxeandia I, et al.** Reporting of methodologic information on trial registries for quality assessment: a study of trial records retrieved from the WHO search portal. PLoS One. 2010;5:e12484. [PMID: 20824212]
48. **Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR.** Comparison of protocols and registry entries to published reports for randomised controlled trials. Cochrane Database Syst Rev. 2011:MR000031. [PMID: 21249714]
49. **Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al.** Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One. 2008;3:e3081. [PMID: 18769481]
50. **GlaxoSmithKline.** Public disclosure of clinical research. Global Public Policy Issues. October 2011. Accessed at www.gsk.com/policies/GSK-on-disclosure-of-clinical-trial-information.pdf on 1 October 2012.
51. **Turner L, Shamseer L, Altman DG, Schulz KF, Moher D.** Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev. 2012;1:60. [PMID: 23194585]

Current Author Addresses: Dr. Chan: Women's College Research Institute, Women's College Hospital, University of Toronto, 790 Bay Street, Toronto, Ontario M5G 1N8, Canada.
Ms. Tetzlaff: Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, Ontario K1H 8L6, Canada.
Dr. Altman: Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Linton Road, Oxford OX2 6UD, United Kingdom.
Dr. Laupacis: Keenan Research Centre at the Li Ka Shing Knowledge Institute of St. Michael's Hospital, 30 Bond Street, Toronto, Ontario M5B 1W8, Canada.
Drs. Gøtzsche and Hróbjartsson: Nordic Cochrane Centre, Rigshospitalet Department 3343, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark.
Dr. Krleža-Jerić: Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario K1H 8M5, Canada.
Dr. Mann: Division of Medical Ethics and Humanities, University of Utah School of Medicine, 75 South 2000 East #108, Salt Lake City, UT 84132.
Dr. Dickersin: Center for Clinical Trials, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Mail Room W5010, Baltimore, MD 21205.
Dr. Berlin: Janssen Research & Development, Janssen Pharmaceutical Companies of Johnson & Johnson, 1125 Trenton Harbourton Road, Titusville, NJ 08560.
Ms. Doré: UK Medical Research Council Clinical Trials Unit, 125 Kingsway, London WC2B 6NH, United Kingdom.
Dr. Parulekar: NCIC Clinical Trials Group, Cancer Research Institute, Queen's University, 10 Stuart Street, Kingston, Ontario K7L 3N6, Canada.
Dr. Summerskill: *The Lancet*, 32 Jamestown Road, London NW1 7BY, United Kingdom.
Dr. Groves: *BMJ*, BMA House, Tavistock Square, London WC1H 9JP, United Kingdom.
Dr. Schulz: Quantitative Sciences, FHI 360, Research Triangle Park, 2224 East NC Highway 54, Durham, NC 27713.
Dr. Sox: The Dartmouth Institute for Health Policy and Clinical Practice, The Geisel School of Medicine at Dartmouth, HB 7500, Hanover, NH 03755.
Dr. Rockhold: GlaxoSmithKline, 5 Moore Drive, PO Box 13398, Research Triangle Park, NC 27709.
Dr. Rennie: The Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, 3333 California Street, Laurel Heights 265, San Francisco, CA 94143-0936.
Dr. Moher: Clinical Epidemiology Program, Ottawa Hospital Research Institute, Department of Epidemiology and Community Medicine, University of Ottawa, 501 Smyth Road, Ottawa, Ontario K1H 8L6, Canada.

Author Contributions: Conception and design: A.-W. Chan, J.M. Tetzlaff, D.G. Altman, A. Laupacis, P.C. Gøtzsche, K. Krleža-Jerić, A. Hróbjartsson, H. Mann, K. Dickersin, J.A. Berlin, W.R. Parulekar, K.F. Schulz, H.C. Sox, D. Rennie, D. Moher.
Analysis and interpretation of the data: A.-W. Chan, J.M. Tetzlaff, D.G. Altman, A. Laupacis, P.C. Gøtzsche, K. Krleža-Jerić, A. Hróbjartsson, K. Dickersin, C.J. Doré, W.R. Parulekar, T. Groves, K.F. Schulz, F.W. Rockhold, D. Rennie, D. Moher.
Drafting of the article: A.-W. Chan, J.M. Tetzlaff, P.C. Gøtzsche, H. Mann, K. Dickersin, J.A. Berlin, C.J. Doré, W.R. Parulekar, K.F. Schulz.
Critical revision of the article for important intellectual content: A.-W. Chan, J.M. Tetzlaff, D.G. Altman, A. Laupacis, K. Krleža-Jerić, A. Hróbjartsson, H. Mann, K. Dickersin, J.A. Berlin, C.J. Doré, W.R. Parulekar, W.S.M. Summerskill, T. Groves, K.F. Schulz, H.C. Sox, F.W. Rockhold, D. Rennie, D. Moher.
Final approval of the article: A.-W. Chan, J.M. Tetzlaff, D.G. Altman, A. Laupacis, P.C. Gøtzsche, K. Krleža-Jerić, A. Hróbjartsson, H. Mann, K. Dickersin, J.A. Berlin, C.J. Doré, W.R. Parulekar, W.S.M. Summerskill, T. Groves, K.F. Schulz, H.C. Sox, F.W. Rockhold, D. Rennie, D. Moher.
Provision of study materials or patients: K. Krleža-Jerić, K. Dickersin.
Statistical expertise: D.G. Altman, P.C. Gøtzsche, C.J. Doré, K.F. Schulz, F.W. Rockhold.
Obtaining of funding: A.-W. Chan, A. Laupacis, D. Moher.
Administrative, technical, or logistic support: A.-W. Chan, P.C. Gøtzsche, K. Krleža-Jerić.
Collection and assembly of data: A.-W. Chan, J.M. Tetzlaff, P.C. Gøtzsche, A. Hróbjartsson, K. Dickersin, C.J. Doré, W.R. Parulekar, K.F. Schulz.

Appendix Table. **World Health Organization Trial Registration Data Set***

Item	Description
1. Primary registry and trial-identifying number	Name of primary registry and the unique identifier assigned by the primary registry
2. Date of registration in primary registry	Date when the trial was officially registered in the primary registry
3. Secondary identifying numbers	Other identifiers, if any Universal Trial Number Identifiers assigned by the sponsor Other trial registration numbers issued by other registries Identifiers issued by funding bodies, collaborative research groups, regulatory authorities, ethics committees/ institutional review boards, etc.
4. Sources of monetary or material support	Major sources of monetary or material support for the trial (e.g., funding agency, foundation, company, institution)
5. Primary sponsor	Person, organization, group, or other legal entity that takes responsibility for initiating and managing a study
6. Secondary sponsor(s)	Additional persons, organizations, or other legal persons, if any, who have agreed with the primary sponsor to take on responsibilities of sponsorship
7. Contact for public queries	E-mail address, telephone number, and postal address of the contact who will respond to general queries, including information about current recruitment status
8. Contact for scientific queries	Name and title, e-mail address, telephone number, postal address, and affiliation of the principal investigator and e-mail address, telephone number, postal address, and affiliation of the contact for scientific queries about the trial (if applicable)
9. Public title	Title intended for the lay public in easily understood language
10. Scientific title	Scientific title of the study as it appears in the protocol submitted for funding and ethical review; include trial acronym, if available
11. Countries of recruitment	Countries from which participants will be recruited
12. Health condition(s) or problem(s) studied	Primary health condition(s) or problem(s) studied (e.g., depression, breast cancer, medication error)
13. Intervention(s)	For each group of the trial, record a brief intervention name plus an intervention description Intervention name: For drugs, use the generic name; for other types of interventions, provide a brief descriptive name Intervention description: Must be sufficiently detailed for it to be possible to distinguish between the groups of a study; for example, interventions involving drugs may include dosage form, dosage, frequency, and duration
14. Key inclusion and exclusion criteria	Inclusion and exclusion criteria for participant selection, including age and sex
15. Study type	Method of allocation (randomized/nonrandomized) Blinding/masking (identify who is blinded) Assignment (e.g., single group, parallel, crossover, factorial) Purpose Phase (if applicable) For randomized trials: Method of sequence generation and allocation concealment
16. Date of first enrollment	Anticipated or actual date of enrollment of the first participant
17. Target sample size	Total number of participants to enroll
18. Recruitment status	Pending: Participants are not yet being recruited or enrolled at any site Recruiting Suspended: Temporary halt in recruitment and enrollment Complete: Participants are no longer being recruited or enrolled Other
19. Primary outcome(s)	The primary outcome should be the outcome used in sample size calculations or the main outcome used to determine the effects of the intervention For each primary outcome provide: Name of the outcome (do not use abbreviations) Metric or method of measurement used (be as specific as possible) Time point of primary interest
20. Key secondary outcome(s)	As for primary outcomes, for each secondary outcome provide: Name of the outcome (do not use abbreviations) Metric or method of measurement used (be as specific as possible) Time point of interest

* Adapted from www.who.int/ictrp/network/trds/en/index.html.

SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials

An-Wen Chan,[1] Jennifer M Tetzlaff,[2] Peter C Gøtzsche,[3] Douglas G Altman,[4] Howard Mann,[5] Jesse A Berlin,[6] Kay Dickersin,[7] Asbjørn Hróbjartsson,[3] Kenneth F Schulz,[8] Wendy R Parulekar,[9] Karmela Krleža-Jeric,[10] Andreas Laupacis,[11] David Moher[2 10]

[1]Women's College Research Institute at Women's College Hospital, Department of Medicine, University of Toronto, Toronto, Canada, M5G 1N8
[2]Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada
[3]Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
[4]Centre for Statistics in Medicine, University of Oxford, Oxford, UK
[5]Division of Medical Ethics and Humanities, University of Utah School of Medicine, Salt Lake City, USA
[6]Janssen Research and Development, Titusville, USA
[7]Center for Clinical Trials, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
[8]Quantitative Sciences, FHI 360, Research Triangle Park, USA
[9]NCIC Clinical Trials Group, Cancer Research Institute, Queen's University, Kingston, Canada
[10]Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada
[11]Keenan Research Centre at the Li Ka Shing Knowledge Institute of St Michael's Hospital, Faculty of Medicine, University of Toronto, Toronto, Canada
Correspondence to: A-W Chan
anwen.chan@utoronto.ca
Accepted: 04 October 2012
Cite this as: *BMJ* 2013;346:e7586
doi: 10.1136/bmj.e7586

High quality protocols facilitate proper conduct, reporting, and external review of clinical trials. However, the completeness of trial protocols is often inadequate. To help improve the content and quality of protocols, an international group of stakeholders developed the SPIRIT 2013 Statement (Standard Protocol Items: Recommendations for Interventional Trials). The SPIRIT Statement provides guidance in the form of a checklist of recommended items to include in a clinical trial protocol.

This SPIRIT 2013 Explanation and Elaboration paper provides important information to promote full understanding of the checklist recommendations. For each checklist item, we provide a rationale and detailed description; a model example from an actual protocol; and relevant references supporting its importance. We strongly recommend that this explanatory paper be used in conjunction with the SPIRIT Statement. A website of resources is also available (www.spirit-statement.org).

The SPIRIT 2013 Explanation and Elaboration paper, together with the Statement, should help with the drafting of trial protocols. Complete documentation of key trial elements can facilitate transparency and protocol review for the benefit of all stakeholders.

Every clinical trial should be based on a protocol—a document that details the study rationale, proposed methods, organisation, and ethical considerations.[1] Trial investigators and staff use protocols to document plans for study conduct at all stages from participant recruitment to results dissemination. Funding agencies, research ethics committees/institutional review boards, regulatory agencies, medical journals, systematic reviewers, and other groups rely on protocols to appraise the conduct and reporting of clinical trials.

To meet the needs of these diverse stakeholders, protocols should adequately address key trial elements. However, protocols often lack information on important concepts relating to study design and dissemination plans.[2-12] Guidelines for writing protocols can help improve their completeness, but existing guidelines vary extensively in their content and have limitations, including non-systematic methods of development, limited stakeholder involvement, and lack of citation of empirical evidence to support their recommendations.[13] As a result, there is also variation in the precise definition and scope of a trial protocol, particularly in terms of its relation to other documents such as procedure manuals.[14]

Given the importance of trial protocols, an international group of stakeholders launched the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) Initiative in 2007 with the primary aim of improving the content of trial protocols. The main outputs are the SPIRIT 2013 Statement,[14] consisting of a 33 item checklist of minimum recommended protocol items (table 1) plus a diagram (fig1); and this accompanying Explanation and Elaboration (E&E) paper. Additional information and resources are also available on the SPIRIT website (www.spirit-statement.org).

The SPIRIT 2013 Statement and E&E paper reflect the collaboration and input of 115 contributors, including trial investigators, healthcare professionals, methodologists, statisticians, trial coordinators, journal editors, as well as representatives from research ethics committees, industry and non-industry funders, and regulatory agencies. Details of the scope and methods have been published elsewhere.[13-15] Briefly, three complementary methods were specified beforehand, in line with current recommendations for development of reporting guidelines[16]: 1) a Delphi consensus survey[15]; 2) two systematic reviews to identify existing protocol guidelines and empirical evidence supporting the importance of specific checklist items; and 3) two face-to-face consensus meetings to finalise the SPIRIT 2013 checklist. Furthermore, the checklist was pilot tested by graduate course students, and an implementation strategy was developed at a stakeholder meeting.

The SPIRIT recommendations are intended as a guide for those preparing the full protocol for a clinical trial. A clinical trial is a prospective study in which one or more

	STUDY PERIOD							
	Enrolment	Allocation	Post-allocation					Close-out
TIMEPOINT*	$-t_1$	0	t_1	t_2	t_3	t_4	*etc*	t_x
ENROLMENT:								
Eligibility screen	X							
Informed consent	X							
(List other procedures)	X							
Allocation		X						
INTERVENTIONS:								
(Intervention A)			←	—	→			
(Intervention B)			X		X			
(List other study groups)			←	—	—	—	→	
ASSESSMENTS:								
(List baseline variables)	X	X						
(List outcome variables)				X		X	etc	X
(List other data variables)			X	X	X	X	etc	X

* List specific timepoints in this row

Fig 1 | Example template for the schedule of enrolment, interventions, and assessments (recommended content can be displayed using other schematic formats). This template is copyrighted by the SPIRIT Group and is reproduced by BMJ with their permission.

interventions are assigned to human participants in order to assess the effects on health related outcomes. The recommendations are not intended to prescribe how a trial should be designed or conducted. Rather, we call for a transparent and complete description of what is intended, regardless of the characteristics or quality of the plans. The SPIRIT 2013 Statement addresses the minimum content for interventional trials; additional concepts may be important to describe in protocols for trials of specific designs (eg, crossover trials) or in protocols intended for submission to specific groups (eg, funders, research ethics committees/institutional review boards). If information for a recommended item is not yet available when the protocol is being finalised (eg, funding sources), this should be explicitly stated and the protocol updated as new information is obtained. Formatting conventions such as a table of contents, glossary of non-standard or ambiguous terms (eg, randomisation phase or off-protocol), and list of abbreviations and references will facilitate understanding of the protocol.

Purpose and development of explanation and elaboration paper

Modelled after other reporting guidelines,[17 18] this E&E paper presents each checklist item with at least one model example from an actual protocol, followed by a full explanation of the rationale and main issues to address. This E&E paper provides important information to facilitate full understanding of each checklist item, and is intended to be used in conjunction with the SPIRIT 2013 Statement.[14] These complementary tools serve to inform trial investigators about important issues to consider in the protocol as they relate to trial design, conduct, reporting, and organisation.

To identify examples for each checklist item, we obtained protocols from public websites, journals, trial investigators, and industry sponsors. Model examples were selected to reflect how key elements could be appropriately described in a trial protocol. Some examples illustrate a specific component of a checklist item, while others encompass all key recommendations for an item. Additional examples are also available on the SPIRIT website (www.spirit-statement.org). The availability of examples for all checklist items indicates the feasibility of addressing each recommended item in the main protocol rather than in separate documents.

Examples are quoted verbatim from the trial protocol. Proper names of trial personnel have been abbreviated with italicised initials, and any reference numbers cited in the original quoted text are denoted by [*Reference*] to distinguish them from references cited in this E&E paper.

For each checklist item we also strived to provide references to empirical data supporting its relevance, which we identified through a systematic review conducted to inform the content of the SPIRIT checklist. We searched MEDLINE, the Cochrane Methodology Register, and the Cochrane Database of Systematic Reviews (limited to methodology reviews) up to September 2009, and EMBASE up to August 2007. We searched reference lists, PubMed "related articles," and citation searches using SCOPUS to identify additional relevant studies. We used piloted forms to screen and extract data relevant to specific checklist items.

Studies were included if they provided empirical data to support or refute the importance of a given protocol concept. A summary of the relevant methodological articles was provided to each E&E author for use in preparing the initial draft text for up to six checklist items; each draft was also reviewed and revised by a second author. When citing empirical evidence in the E&E, we aimed to reference a systematic review when available. When no review was identified, we either cited all relevant individual studies, or if too numerous, a representative sample of the literature. Some items had little or no identified empirical evidence (eg, title) but their inclusion in the checklist is supported by a strong pragmatic or ethical rationale. Where relevant, we also provide references to non-empirical publications for further reading.

Two lead authors (AWC, JMT) collated and refined the content and format for all items, and then circulated three iterations of an overall draft to the coauthors for editing and final approval.

SPIRIT 2013 Explanation and Elaboration

Section 1: Administrative information

Item 1: Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym

Example

"A multi-center, investigator-blinded, randomized, 12-month, parallel-group, non-inferiority study to compare the efficacy of 1.6 to 2.4 g Asacol® Therapy QD [once daily] versus divided dose (BID [twice daily]) in the maintenance of remission of ulcerative colitis."[19]

Explanation

The title provides an important means of trial identification. A succinct description that conveys the topic (study population, interventions), acronym (if any), and basic study design—including the method of intervention allocation (eg, parallel group randomised trial; single-group trial)—will facilitate retrieval from literature or internet

Table 1 | SPIRIT 2013 checklist: recommended items to address in a clinical trial protocol and related documents*

Section/item	ItemNo	Description
Administrative information		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry
	2b	All items from the World Health Organization Trial Registration Data Set
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material, and other support
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors
	5b	Name and contact information for the trial sponsor
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)
Introduction		
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention
	6b	Explanation for choice of comparators
Objectives	7	Specific objectives or hypotheses
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)
Methods: Participants, interventions, and outcomes		
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see fig 1)
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size
Methods: Assignment of interventions (for controlled trials)		
Allocation:		
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts) and how
	17b	If blinded, circumstances under which unblinding is permissible and procedure for revealing a participant's allocated intervention during the trial
Methods: Data collection, management, and analysis		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)
Methods: Monitoring		
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed

Continued

Section/item	ItemNo	Description
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor
Ethics and dissemination		
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions
	31b	Authorship eligibility guidelines and any intended use of professional writers
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code
Appendices		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable

*Amendments to the protocol should be tracked and dated. The SPIRIT checklist belongs to the SPIRIT Group and is reproduced by BMJ with their permission

searches and rapid judgment of relevance.[20] It can also be helpful to include the trial framework (eg, superiority, non-inferiority), study objective or primary outcome, and if relevant, the study phase (eg, phase II).

Trial registration—registry

Item 2a: Trial identifier and registry name. If not yet registered, name of intended registry

Example

"EudraCT: 2010-019180-10
ClinicalTrials.gov: NCT01066572
ISRCTN: 54540667."[21]

Explanation

There are compelling ethical and scientific reasons for trial registration.[22-24] Documentation of a trial's existence on a publicly accessible registry can help to increase transparency,[24 25] decrease unnecessary duplication of research effort, facilitate identification of ongoing trials for prospective participants, and identify selective reporting of study results.[26-28] As mandated by the International Committee of Medical Journal Editors (ICMJE) and jurisdictional legislation,[29-31] registration of clinical trials should occur before recruitment of the first trial participant.

We recommend that registry names and trial identifiers assigned by the registries be prominently placed in the protocol, such as on the cover page. If the trial is not yet registered, the intended registry should be indicated and the protocol updated upon registration. When registration in multiple registries is required (eg, to meet local regulation), each identifier should be clearly listed in the protocol and each registry.

Trial registration—data set

Item 2b: All items from the World Health Organization Trial Registration Data Set

Example: see table 2

Explanation

In addition to a trial registration number, the World Health Organization (WHO) recommends a minimum standard list of items to be included in a trial registry in order for a trial to be considered fully registered (www.who.int/ictrp/network/trds/en/index.html). These standards are supported by ICMJE, other journal editors, and jurisdictional legislation.[29-31] We recommend that the WHO Trial Registration Data Set be included in the protocol to serve as a brief structured summary of the trial. Its inclusion in the protocol can also signal updates for the registry when associated protocol sections are amended—thereby promoting consistency between information in the protocol and registry.

Protocol version

Item 3: Date and version identifier

Example

"Issue date: 25 Jul 2005
Protocol amendment number: 05
Authors: *MD,JH*

Revision chronology:

UM . . . 00, 2004-Jan-30 Original
UM . . . 01, 2004-Feb-7 Amendment 01.:
Primary reason for amendment: changes in Section 7.1 regarding composition of comparator placebo
Additional changes (these changes in and of themselves would not justify a protocol amendment): correction of typographical error in Section 3.3 . . .
UM . . . 05, 2005-Jul-25 Amendment No.5:
At the request of US FDA statements were added to the protocol to better clarify and define the algorithm for determining clinical or microbiological failures prior to the follow-up visit."[33]

Explanation

Sequentially labelling and dating each protocol version helps to mitigate potential confusion over which document

Table 2 | Example of trial registration data

Data category	Information[32]
Primary registry and trial identifying number	ClinicalTrials.gov NCT01143272
Date of registration in primary registry	11 June, 2010
Secondary identifying numbers	BNI-2009-01, 2009-017374-20, ISRCTN01005546, DRKS00000084
Source(s) of monetary or material support	Bernhard Nocht Institute for Tropical Medicine
Primary sponsor	Bernhard Nocht Institute for Tropical Medicine
Secondary sponsor(s)	German Federal Ministry of Education and Research
Contact for public queries	SE, MD, MPH [email address]
Contact for scientific queries	SE, MD, MPH Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany
Public title	Probiotic *Saccharomyces boulardii* for the prevention of antibiotic associated diarrhoea (SacBo)
Scientific title	*S boulardii* for the prevention of antibiotic associated diarrhoea—randomised, double blind, placebo controlled trial
Countries of recruitment	Germany
Health condition(s) or problem(s) studied	Antibiotic treatment, *Clostridium difficile*, diarrhoea
Intervention(s)	Active comparator: *S boulardii* (500 mg *S boulardii* per day)
	Placebo comparator: microcristallin cellulose (matching capsules containing no active ingredients)
Key inclusion and exclusion criteria	Ages eligible for study: ≥18 years; Sexes eligible for study: both; Accepts healthy volunteers: no
	Inclusion criteria: adult patient (≥ 18 years), patient hospitalised . . .
	Exclusion criteria: allergy against yeast and/or Perenterol forte and/or placebos containing *S cerevisiae* HANSEN CBS 5926, lactose monohydrate, magnesium stearate, gelatine, sodium dodecyl sulfate, titan dioxide, microcrystalline cellulose
Study type	Interventional
	Allocation: randomized; Intervention model: parallel assignment; Masking: double blind . . .
	Primary purpose: prevention
	Phase III
Date of first enrolment	June 2010
Target sample size	1520
Recruitment status	Recruiting
Primary outcome(s)	Cumulative incidence of any antibiotic associated diarrhoea (time frame: 2 years; not designated as safety issue)
Key secondary outcomes	Cumulative incidence of *C difficile* associated diarrhoea (time frame: 2 years; not designated as safety issue) . . .

is the most recent. Explicitly listing the changes made relative to the previous protocol version is also important (see Item 25). Transparent tracking of versions and amendments facilitates trial conduct, review, and oversight.

Funding

Item 4: Sources and types of financial, material, and other support

Example

"Tranexamic acid will be manufactured by Pharmacia (Pfizer, Sandwich, UK) and placebo by South Devon Healthcare NHS Trust, UK. The treatment packs will be prepared by an independent clinical trial supply company (Brecon Pharmaceuticals Limited, Hereford, UK) . . .
LSHTM [London School of Hygiene and Tropical Medicine] is funding the run-in costs for the WOMAN trial and up to 2,000 patients' recruitment. The main phase is funded by the UK Department of Health and the Wellcome Trust. Funding for this trial covers meetings and central organisational costs only. Pfizer, the manufacturer of tranexamic acid, have provided the funding for the trial drug and placebo used for this trial. An educational grant, equipment and consumables for ROTEM [thromboelastometry procedure] analysis has been provided by Tem Innovations GmbH, M.-Kollar-Str. 13-15, 81829 Munich, Germany for use in the WOMAN-ETAC study. An application for funding to support local organisational costs has been made to University of Ibadan Senate Research Grant. The design, management, analysis and reporting of the study are entirely independent of the manufacturers of tranexamic acid and Tem Innovations GmbH."[34]

Explanation

A description of the sources of financial and non-financial support provides relevant information to assess study feasibility and potential competing interests (Item 28). Although both industry funded and non-industry funded trials are susceptible to bias,[4 35] the former are more likely to report trial results and conclusions that favour their own interventions.[27 36-39] This tendency could be due to industry trials being more likely to select effective interventions for evaluation (Item 6a), to use less effective control interventions (Item 6b), or to selectively report outcomes (Item 12), analyses (Item 20) or full studies (Item 31).[38 40-43] Non-financial support (eg, provision of drugs) from industry has not been shown to be associated with biased results, although few studies have examined this issue.[44 45]

At a minimum, the protocol should identify the sources of financial and non-financial support; the specific type (eg, funds, equipment, drugs, services) and time period of support; and any vested interest that the funder may have in the trial. If a trial is not yet funded when the protocol is first written, the proposed sources of support should be listed and updated as funders are confirmed.

No clear consensus exists regarding the level of additional funding details that should be provided in the trial protocol as opposed to trial contracts, although full disclosure of funding information in the protocol can help to better identify financial competing interests. Some jurisdictional guidelines require more detailed disclosure, including monetary amounts granted from each funder, the mechanism of providing financial support (eg, paid in fixed sum or per recruited participant), and the specific fund recipient (eg, trial investigator, department/institute).[46] Detailed disclosure allows research ethics committees/institutional review boards (REC/IRBs) to assess whether the reimbursement amount is reasonable in relation to the time and expenses incurred for trial conduct.

Roles and responsibilities—contributorship

Item 5a: Names, affiliations, and roles of protocol contributors

Example

"*RTL* [address], *EJM* [address], *AK* [address] . . .

Authors' contributions

RTL conceived of the study. *AK*, *EN*, *SB*, *PR*, *WJ*, *JH*, and *MC* initiated the study design and *JK* and *LG* helped with implementation. *RTL*, *JK*, *LG*, and *FP* are grant holders. *LT* and *EM* provided statistical expertise in clinical trial design and *RN* is conducting the primary statistical analysis. All authors contributed to refinement of the study protocol and approved the final manuscript."[47]

Explanation

Individuals who contribute substantively to protocol development and drafting should have their contributions reported. As with authorship of journal articles,[48] listing the protocol contributors, their affiliations, and their roles in the protocol development process provides due recognition, accountability, and transparency. Naming of contributors can also help to identify competing interests and reduce ghost authorship (Items 28 and 31b).[9 10] If professional medical writers are employed to draft the protocol, then this should be acknowledged as well.

Naming of authors and statements of contributorship are standard for protocols published in journals such as *Trials*[49] but are uncommon for unpublished protocols. Only five of 44 industry-initiated protocols approved in 1994-95 by a Danish research ethics committee explicitly identified the protocol authors.[9]

Roles and responsibilities—sponsor contact information

Item 5b: Name and contact information for the trial sponsor

Example

"Trial Sponsor: University of Nottingham
Sponsor's Reference: RIS 8024 . . .
Contact name: Mr *PC*
Address: King's Meadow Campus . . .
Telephone: . . .
Email: . . ."[50]

Explanation

The sponsor can be defined as the individual, company, institution, or organisation assuming overall responsibility for the initiation and management of the trial, and is not necessarily the main funder.[51 52] In general, the company is the sponsor in industry initiated trials, while the funding agency or institution of the principal investigator is often the sponsor for investigator initiated trials. For some investigator initiated trials, the principal investigator can be considered to be a "sponsor-investigator" who assumes both sponsor and investigator roles.[51 53]

Identification of the trial sponsor provides transparency and accountability. The protocol should identify the name, contact information, and if applicable, the regulatory agency identifying number of the sponsor.

Roles and responsibilities—sponsor and funder

Item 5c: Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities

Example

"This funding source had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision to submit results."[54]

Explanation

There is potential for bias when the trial sponsor or funder (sometimes the same entity) has competing interests (Item 28) and substantial influence on the planning, conduct, or reporting of a trial. Empirical research indicates that specific forms of bias tend to be more prevalent in trials funded by industry compared to those funded by non-commercial sources.[36-38 45 55-60] The design, analysis, interpretation, and reporting of most industry-initiated trials are controlled by the sponsor; this authority is often enforced by contractual agreements signed between the sponsor and trial investigators (Item 29).[10 61]

The protocol should explicitly outline the roles and responsibilities of the sponsor and any funders in study design, conduct, data analysis and interpretation, manuscript writing, and dissemination of results. It is also important to state whether the sponsor or funder controls the final decision regarding any of these aspects of the trial.

Despite the importance of declaring the roles of the trial sponsor and funders, few protocols explicitly do so. Among 44 protocols for industry-initiated trials receiving ethics approval in Denmark from 1994-95, none stated explicitly who had contributed to the design of the trial.[9]

Roles and responsibilities—committees

Item 5d: Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)

Explanation

The protocol should outline the general membership of the various committees or groups involved in trial coordination and conduct; describe the roles and responsibilities of each; and (when known) identify the chairs and members. This information helps to ensure that roles and responsibilities are clearly understood at the trial onset, and facilitates communication from external parties regarding the trial. It also enables readers to understand the mandate and expertise of those responsible for overseeing participant safety, study design, database integrity, and study conduct. For example, empirical evidence supports the pivotal role of an epidemiologist or biostatistician in designing and conducting higher quality trials.[63 64]

Example

"Principal investigator and research physician
Design and conduct of RITUXVAS
Preparation of protocol and revisions
Preparation of investigators brochure (IB) and CRFs [case report forms]
Organising steering committee meetings
Managing CTO [clinical trials office]
Publication of study reports
Members of TMC [Trial Management Committee]

Steering committee (SC)
(see title page for members)
Agreement of final protocol
All lead investigators will be steering committee members. One lead investigator per country will be nominated as national coordinator.
Recruitment of patients and liaising with principle [sic] investigator
Reviewing progress of study and if necessary agreeing changes to the protocol and/or investigators brochure to facilitate the smooth running of the study.

Trial management committee (TMC)
(Principle [sic] investigator, research physician, administrator)
Study planning
Organisation of steering committee meetings
Provide annual risk report MHRA [Medicines and Healthcare Products Regulatory Agency] and ethics committee
SUSAR [Serious unexpected suspected adverse events] reporting to MHRA and Roche
Responsible for trial master file
Budget administration and contractual issues with individual centres
Advice for lead investigators
Audit of 6 monthly feedback forms and decide when site visit to occur.
Assistance with international review, board/independent ethics committee applications
Data verification
Randomisation
Organisation of central serum sample collection

Data manager
Maintenance of trial IT system and data entry
Data verification

Lead investigators
In each participating centre a lead investigator (senior nephrologist/rheumatologist/ immunologist) will be identified, to be responsible for identification, recruitment, data collection and completion of CRFs, along with follow up of study patients and adherence to study protocol and investigators brochure. . . . Lead investigators will be steering committee members, with one investigator per country being nominated as national coordinator."[62]

Example

"Background
Introduction: For people at ages 5 to 45 years, trauma is second only to HIV/AIDS as a cause of death. . . .
Mechanisms: The haemostatic system helps to maintain the integrity of the circulatory system after severe vascular injury, whether traumatic or surgical in origin.[reference] Major surgery and trauma trigger similar haemostatic responses . . . Antifibrinolytic agents have been shown to reduce blood loss in patients with both normal and exaggerated fibrinolytic responses to surgery, and do so without apparently increasing the risk of post-operative complications, . . .
Existing knowledge: Systemic antifibrinolytic agents are widely used in major surgery to prevent fibrinolysis and thus reduce surgical blood loss. A recent systematic review [reference] of randomised controlled trials of antifibrinolytic agents (mainly aprotinin or tranexamic acid) in elective surgical patients identified 89 trials including 8,580 randomised patients (74 trials in cardiac, eight in orthopaedic, four in liver, and three in vascular surgery). The results showed that these treatments reduced the numbers needing transfusion by one third, reduced the volume needed per transfusion by one unit, and halved the need for further surgery to control bleeding. These differences were all highly statistically significant. There was also a statistically non-significant reduction in the risk of death (RR=0.85: 95% CI 0.63 to 1.14) in the antifibrinolytic treated group.
. . .
Need for a trial: A simple and widely practicable treatment that reduces blood loss following trauma might prevent thousands of premature trauma deaths each year and secondly could reduce exposure to the risks of blood transfusion. Blood is a scarce and expensive resource and major concerns remain about the risk of transfusion-transmitted infection. . . . A large randomised trial is therefore needed of the use of a simple, inexpensive, widely practicable antifibrinolytic treatment such as tranexamic acid . . . in a wide range of trauma patients who, when they reach hospital are thought to be at risk of major haemorrhage that could significantly affect their chances of survival.

Dose selection
The systematic review of randomised controlled trials of antifibrinolytic agents in surgery showed that dose regimens of tranexamic acid vary widely.[reference] . . .
In this emergency situation, administration of a fixed dose would be more practicable as determining the weight of a patient would be impossible. Therefore a fixed dose within the dose range which has been shown to inhibit fibrinolysis and provide haemostatic benefit is being used for this trial. . . . The planned duration of administration allows for the full effect of tranexamic acid on the immediate risk of haemorrhage without extending too far into the acute phase response seen after surgery and trauma."[65]

Section 2: Introduction

Background and rationale

Item 6a: Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention

Explanation

The value of a research question, as well as the ethical and scientific justification for a trial, depend to a large degree on the uncertainty of the comparative benefits or harms of the interventions, which depends in turn on the existing body of knowledge on the topic. The background section of a protocol should summarise the importance of the research question, justify the need for the trial in the context of available evidence, and present any available data regarding the potential effects of the interventions (efficacy and harms).[66 67] This information is particularly important to the trial participants and personnel, as it provides motivation for contributing to the trial.[68 69] It is also relevant to funders, REC/IRBs, and other stakeholders who evaluate the scientific and ethical basis for trial conduct.

To place the trial in the context of available evidence, it is strongly recommended that an up-to-date systematic review of relevant studies be summarised and cited in the protocol.[70] Several funders request this information in grant applications.[71 72] Failure to review the cumulated evidence can lead to unnecessary duplication of research or to trial participants being deprived of effective, or exposed to harmful, interventions.[73-76] A minority of published trial reports cite a systematic review of pre-existing evidence,[77 78] and in one survey only half of trial investigators

were aware of a relevant existing review when they had designed their trial.[79] Given that about half of trials remain unpublished,[80-82] and that published trials often represent a biased subset of all trials,[80 83] it is important that systematic reviews include a search of online resources such as trial registries, results databases, and regulatory agency websites.[84]

Background and rationale—choice of comparators

Item 6b: Explanation for choice of comparators

Example

"Choice of comparator

In spite of the increasing numbers of resistant strains, chloroquine monotherapy is still recommended as standard blood-stage therapy for patients with *P* [*Plasmodium*] *vivax* malaria in the countries in which this trial will be conducted. Its selection as comparator is therefore justified. The adult dose of chloroquine will be 620 mg for 2 days followed by 310 mg on the third day and for children 10 mg/kg for the first two days and 5 mg/kg for the third day. Total dose is in accordance with the current practice in the countries where the study is conducted. The safety profile of chloroquine is well established and known. Although generally well tolerated, the following side-effects of chloroquine treatment have been described: Gastro-intestinal disturbances, headache, hypotension, convulsions, visual disturbances, depigmentation or loss of hair, skin reactions (rashes, pruritus) and, rarely, bone-marrow suppression and hypersensitivity reactions such as urticaria and angioedema. Their occurrence during the present trial may however be unlikely given the short (3-day) duration of treatment."[85]

Explanation

The choice of control interventions has important implications for trial ethics, recruitment, results, and interpretation. In trials comparing an intervention to an active control or usual care, a clear description of the rationale for the comparator intervention will facilitate understanding of its appropriateness.[86 87] For example, a trial in which the control group receives an inappropriately low dose of an active drug will overestimate the relative efficacy of the study intervention in clinical practice; conversely, an inappropriately high dose in the control group will lead to an underestimate of the relative harms of the study intervention.[87 88]

The appropriateness of using placebo-only control groups has been the subject of extensive debate and merits careful consideration of the existence of other effective treatments, the potential risks to trial participants, and the need for assay sensitivity—that is, ability to distinguish an effective intervention from less effective or ineffective interventions.[89 90] In addition, surveys have demonstrated that a potential barrier to trial participation is the possibility of being allocated a placebo-only or active control intervention that is perceived to be less desirable than the study intervention.[68 69 91 92] Evidence also suggests that enrolled participants perceive the effect of a given intervention differently depending on whether the control group consists of an active comparator or only placebo.[93-96]

Finally, studies suggest that some "active" comparators in head-to-head randomised trials are presumed by trial investigators to be effective despite having never previously been shown to be superior to placebo.[74 97] In a systematic review of over 100 head-to-head antibiotic trials for mild to moderate chronic obstructive pulmonary disease,[74] cumulative meta-analysis of preceding placebo controlled trials did not show a significant effect of antibiotics over placebo. Such studies again highlight the importance of providing a thorough background and rationale for a trial and the choice of comparators—including data from an up-to-date systematic review—to enable potential participants, physicians, REC/IRBs, and funders to discern the merit of the trial.

Objectives

Item 7: Specific objectives or hypotheses

Example

"1.1 Research hypothesis

Apixaban is noninferior to warfarin for prevention of stroke (hemorrhagic, ischemic or of unspecified type) or systemic embolism in subjects with atrial fibrillation (AF) and additional risk factor(s) for stroke.

...

2 STUDY OBJECTIVES

2.1 Primary objective

To determine if apixaban is noninferior to warfarin (INR [international normalized ratio] target range 2.0-3.0) in the combined endpoint of stroke (hemorrhagic, ischemic or of unspecified type) and systemic embolism, in subjects with AF and at least one additional risk factor for stroke.

2.2 Secondary objectives

2.2.1 Key secondary objectives

The key secondary objectives are to determine, in subjects with AF and at least one additional risk factor for stroke, if apixaban is superior to warfarin (INR target range 2.0 - 3.0) for,

- the combined endpoint of stroke (hemorrhagic, ischemic or of unspecified type) and systemic embolism
- major bleeding [International Society of Thrombosis and Hemostasis]
- all-cause death

2.2.2 Other secondary objectives

- To compare, in subjects with AF and at least one additional risk factor for stroke, apixaban and warfarin with respect to:
  - The composite endpoint of stroke (ischemic, hemorrhagic, or of unspecified type), systemic embolism and major bleeding, in warfarin naive subjects

...

- To assess the safety of apixaban in subjects with AF and at least one additional risk factor for stroke."[98]

Explanation

The study objectives reflect the scientific questions to be answered by the trial, and define its purpose and scope. They are closely tied to the trial design (Item 8) and analysis methods (Item 20). For example, the sample size calculation and statistical analyses for superiority trials will differ from those investigating non-inferiority.

The objectives are generally phrased using neutral wording (eg, "to compare the effect of treatment A versus treatment B on outcome X") rather than in terms of a particular direction of effect.[99] A hypothesis states the predicted effect of the interventions on the trial outcomes. For multiarm trials, the objectives should clarify the way in which all the treatment groups will be compared (eg, A versus B; A versus C).

Trial design

Item 8: Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, non-inferiority, exploratory)

Example

"The PROUD trial is designed as a randomised, controlled, observer, surgeon and patient blinded multicenter superiority trial with two parallel groups and a primary endpoint of wound infection during 30 days after surgery … randomization will be performed as block randomization with a 1:1 allocation."[100]

Explanation

The most common design for published randomised trials is the parallel group, two arm, superiority trial with 1:1 allocation ratio.[101] Other trial types include crossover, cluster, factorial, split body, and n of 1 randomised trials, as well as single group trials and non-randomised comparative trials.

For trials with more than one study group, the allocation ratio reflects the intended relative number of participants in each group (eg, 1:1 or 2:1). Unequal allocation ratios are used for a variety of reasons, including potential cost savings, allowance for learning curves, and ethical considerations when the balance of existing evidence appears to be in favour of one intervention over the other.[102] Evidence also suggests a preference of some participants for enrolling in trials with an allocation ratio that favours allocation to an active treatment.[92]

The framework of a trial refers to its overall objective to test the superiority, non-inferiority, or equivalence of one intervention with another, or in the case of exploratory pilot trials, to gather preliminary information on the intervention (eg, harms, pharmacokinetics) and the feasibility of conducting a full-scale trial.

It is important to specify and explain the choice of study design because of its close relation to the trial objectives (Item 7) and its influence on the study methods, conduct, costs,[103] results,[104-106] and interpretation. For example, factorial and non-inferiority trials can involve more complex methods, analyses, and interpretations than parallel group superiority trials.[107 108] In addition, the interpretation of trial results in published reports is not always consistent with the prespecified trial framework,[6 109 110] especially among reports claiming post hoc equivalence based on a failure to demonstrate superiority rather than a specific test of equivalence.[109]

There is increasing interest in adaptive designs for clinical trials, defined as the use of accumulating data to decide how to modify aspects of a study as it continues, without undermining the validity and integrity of the trial.[111 112] Examples of potential adaptations include stopping the trial early, modifying the allocation ratio, re-estimating the sample size, and changing the eligibility criteria. The most valid adaptive designs are those in which the opportunity to make adaptations is based on prespecified decision rules that are fully documented in the protocol (Item 21b).

Section 3a: Methods—participants, interventions, and outcomes

Study setting

Item 9: Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained

Example

"Selection of countries

…To detect an intervention-related difference in HIV incidences with the desired power, the baseline incidences at the sites must be sufficiently high. We chose the participating sites so that the average baseline annual incidence across all communities in the study is likely to reach at least 3%. The various sites in sub-Saharan Africa met this criterion, but we also wanted sites in Asia to extend the generalizability of the intervention. The only location in Asia with sufficient incidence at the community level is in ethnic minority communities in Northern Thailand, where HIV incidence is currently in excess of 7%;[reference] thus they were invited to participate as well. Our final selection of sites combines rural (Tanzania, Zimbabwe, Thailand, and KwaZulu-Natal) and an urban (Soweto) location. The cultural circumstances between the sub-Saharan African sites vary widely…

Definition of community

Each of the three southern African sites (Harare, Zimbabwe; and Soweto and Vulindlela, South Africa) selected eight communities, the East African (Tanzanian) site selected 10 communities, and Thailand selected 14 communities…They are of a population size of approximately 10,000… which fosters social familiarity and connectedness, and they are geographically distinct. Communities are defined primarily geographically for operational purposes for the study, taking into account these dimensions of social communality. The communities chosen within each country and site are selected to be sufficiently distant from each other so that there would be little cross-contamination or little possibility that individuals from a control community would benefit from the activities in the intervention community."[113]

Explanation

A description of the environment in which a trial will be conducted provides important context in terms of the applicability of the study results; the existence and type of applicable local regulation and ethics oversight; and the type of healthcare and research infrastructure available. These considerations can vary substantially within and between countries.

At a minimum, the countries, type of setting (eg, urban versus rural), and the likely number of study sites should be reported in the protocol. These factors have been associated with recruitment success and degree of attrition for some trials,[68 91 92 114-117] but not for others.[118 119] Trial location has also been associated with trial outcome,[120] aspects of trial quality (eg, authenticity of randomisation[121]), and generalisability.[122]

Eligibility criteria

Item 10: Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)

Explanation

Eligibility criteria for potential trial participants define the study population. They can relate to demographic information; type or severity of the health condition; comorbidities; previous or current treatment; diagnostic procedures; pregnancy; or other relevant considerations.[125] In trials of operator-dependent interventions such as surgery and psychotherapy, it is usually important to promote consistency of intervention delivery by also defining the eligibility criteria for care providers and centres where the intervention will be administered.[126]

Clear delineation of eligibility criteria serves several purposes. It enables study personnel to apply these criteria consistently throughout the trial.[127] The choice of eligibility criteria can affect recruitment and attri-

Examples

"Patients (or a representative) must provide written, informed consent before any study procedures occur (see Appendix 1 for sample Informed Consent Form) . . .

5.1. Inclusion Criteria

Patients eligible for the trial must comply with all of the following at randomization:

1. Age ≥16 years
2. Current admission under the care of the heart-failure service at the site

. . .

5.2. Exclusion Criteria

1. Acute decompensation thought by the attending heart-failure physician to require or be likely to require PAC [pulmonary-artery catheter] during the next 24 hours. Such patients should be entered into the PAC Registry (see below).
2. Inability to undergo PAC placement within the next 12 hours

. . .

Patients enrolled in other investigational drug studies are potential candidates for ESCAPE. As the ESCAPE protocol does not involve any investigational agents or techniques, patients would be eligible for dual randomization if they are on stable doses of the investigational drugs. . . .

13. Study Network, Training, and Responsibilities

. . . To qualify, physicians responsible for PAC [pulmonary-artery catheter] placements will be required to show proof of insertion of ≥50 PACs in the previous year with a complication rate of <5%. Further, clinicians will need to show competence in the following areas to participate in the study: 1) insertion techniques and cardiovascular anatomy; 2) oxygen dynamics; . . . and 7) common PAC complications. [reference] . . . we will assume basic competence in these areas after satisfactory completion of the PACEP [PAC educational programme] module."[123]

"Trial centre requirements

A number of guidelines have stated thrombolysis should only be considered if the patient is admitted to a specialist centre with appropriate experience and expertise.[reference] Hospitals participating in IST-3 [third International Stroke Trial] should have an organized acute stroke service. The components of effective stroke unit care have been identified . . . In brief, the facilities (details of these requirements are specified in the separate operations manual) should include:

- Written protocol for the acute assessment of patients with suspected acute stroke to include interventions to reduce time from onset to treatment.
- Immediate access to CT [computed tomographic] or MR [magnetic resonance] brain scanning (preferably 24 hours a day).

A treatment area where thrombolysis may be administered and the patient monitored according to trial protocol, preferably an acute stroke unit."[124]

tion,[67 114 115 117 118 128-130] as well as outcome event rates.[39 131] In addition, the criteria convey key information related to external validity (generalisability or applicability).[132] The importance of transparent documentation is highlighted by evidence that the eligibility criteria listed in publications are often different from those specified in the protocol.[125 133 134]

Certain eligibility criteria warrant explicit justification in the protocol, particularly when they limit the trial sample to a narrow subset of the population.[132 135 136] The appropriateness of restrictive participant selection depends on the trial objectives.[137] When trial participants differ substantially from the overall population to whom the intervention will be applied, the trial results may not reflect the impact in real world practice settings.[134 138-144]

Interventions

Item 11a: Interventions for each group with sufficient detail to allow replication, including how and when they will be administered

Explanation

Studies of trials and systematic reviews have shown that important elements of the interventions are not described in half of the publications.[146 147] If such elements are also missing from the protocol, or if the protocol simply refers to other documents that are not freely accessible, then it can be impossible for healthcare providers, systematic reviewers,

Example

"Eligible patients will be randomised in equal proportions between IL-1ra [interleukin-1 receptor antagonist] and placebo, receiving either a once daily, subcutaneous (s.c.) injection of IL-1ra (dose 100 mg per 24 h) for 14 days, or a daily s.c. injection of placebo for 14 days . . .

The study drug and placebo will be provided by Amgen Inc in its commercially available recombinant form . . . The study drug and placebo will be relabelled by Amgen, in collaboration with CTEU [Clinical Trials and Evaluation Unit] according to MHRA [Medicines and Healthcare Products Regulatory Agency] guidelines.

The first dose of IL-1ra will be given within 24 h +2 h of the positive Troponin. Injections will be given at a standardised time (24 ± 2 h after the previous dose), immediately after blood sampling. IL-1ra or placebo will [be] administered to the patient by the research nurse while the patient is in hospital. During the hospital stay, the patient will be taught to self-administer the injection by the research nurse and on discharge will continue at home. This has proven possible in other ACS [acute coronary syndrome] trials that required self injection of subcutaneous heparin [reference]. Full written guidance on self injection will also be provided to patients. If self injection is found not to be possible in an individual patient for unexpected reasons, an alternative method will be sought (eg district nurse, or attending the hospital) to try and maintain full compliance with scheduled study drug regimen after discharge. Patients will also be asked to complete a daily injection diary. All personnel will be blinded to the identity of the syringe contents."[145]

policymakers, and others to fully understand, implement, or evaluate the trial intervention.[148] This principle applies to all types of interventions, but is particularly true for complex interventions (eg, health service delivery; psychotherapy), which consist of interconnected components that can vary between healthcare providers and settings.

For drugs, biological agents, or placebos, the protocol description should include the generic name, manufacturer, constituent components, route of administration, and dosing schedule (including titration and run-in periods, if applicable).[149 150] The description of non-drug interventions—such as devices, procedures, policies, models of care, or counselling—is generally more complex and warrants additional details about the setting (Item 9) and individuals administering the interventions. For example, the level of pre-trial expertise (Item 10) and specific training of individuals administering these complex interventions are often relevant to describe (eg, for surgeons, psychologists, physiotherapists). When intervention delivery is subject to variation, it is important to state whether the same individuals will deliver the trial interventions in all study groups, or whether different individuals will manage each study group—in which case it can be difficult to separate the effect of the intervention from that of the individual delivering it. Interventions that consist of "usual care" or "standard of care" require further elaboration in the protocol, as this care can vary substantially across centres and patients, as well as over the duration of the trial.

Interventions—modifications

Item 11b: Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)

Explanation

For a given trial participant, the assigned study intervention may need to be modified or discontinued by

Example

"Gastro-Intestinal Upset
The tablets may be taken in two equally divided doses, if necessary, to improve gastro-intestinal tolerance. Should it be necessary the daily dose may be reduced by one tablet at a time to improve gastro-intestinal tolerance.

Renal Function Impairment
Since sodium clodronate is excreted unchanged by the kidney its use is contra-indicated in patients with moderate to severe renal impairment (serum creatinine greater than 2 times upper limit of normal range of the centre). If renal function deteriorates to this extent the trial medication should be withdrawn from the patient. This should be reported as an adverse event. In patients with normal renal function or mild renal impairment (serum creatinine less than 2 times upper limit of normal range of the centre) serum creatinine should be monitored during therapy.

Allergic Reactions
Allergic skin reactions have been observed in rare cases. If this is suspected withdraw the trial medication from the patient. This should be reported as an adverse event.

Biochemical Disturbances
Asymptomatic hypocalcaemia has been noted rarely. Temporary suspension of the trial medication until the serum calcium returns into the normal range is recommended. The trial medication can be then restarted at half the previous dose. If the situation returns withdraw the trial medication from the patient. This should be reported as an adverse event…"[151]

trial investigators for various reasons, including harms, improved health status, lack of efficacy, and withdrawal of participant consent. Comparability across study groups can be improved, and subjectivity in care decisions reduced, by defining standard criteria for intervention modifications and discontinuations in the protocol. Regardless of any decision to modify or discontinue their assigned intervention, study participants should be retained in the trial whenever possible to enable follow-up data collection and prevent missing data (Item 18b).[152]

Interventions—adherence

Item 11c: Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return; laboratory tests)

Example

"Adherence reminder sessions
Face-to-face adherence reminder sessions will take place at the initial product dispensing and each study visit thereafter. This session will include:
- The importance of following study guidelines for adherence to once daily study product
- Instructions about taking study pills including dose timing, storage, and importance of taking pills whole, and what to do in the event of a missed dose.
- Instructions about the purpose, use, and care of the MEMS® cap [medication event monitoring system] and bottle
- Notification that there will be a pill count at every study visit
- Reinforcement that study pills may be TDF [tenofovir disproxil fumarate] or placebo
- Importance of calling the clinic if experiencing problems possibly related to study product such as symptoms, lost pills or MEMS® cap.

Subsequent sessions will occur at the follow-up visits. Participants will be asked about any problems they are having taking their study pills or using the MEMS® cap. There will be brief discussion of reasons for missed doses and simple strategies for enhancing adherence, eg, linking pill taking to meals or other daily activities. Participants will have an opportunity to ask questions and key messages from the initial session will be reviewed as needed . . .

Adherence assessments
To enhance validity of data, multiple methods will be used to assess medication adherence including pill count; an electronic medication event monitoring system (MEMS® cap) [reference]; and ACASI [audio-computer administered interview] questionnaire items including a one month visual analogue scale,[reference] reasons for non-compliance, and use of the MEMS® cap. Participants will return the unused tablets and bottle at each follow-up visit. Unused tablets will be counted and recorded on the appropriate CRF [case report form]. Electronic data collected in the MEMS® cap will be downloaded into a designated, secure study computer."[153]

Explanation
Adherence to intervention protocols refers to the degree to which the behaviour of trial participants corresponds to the intervention assigned to them.[154] Distinct but related concepts include trial retention (Item 18b) and adherence to the follow-up protocol of procedures and assessments (Item 13).

On average, adherence to intervention protocols is higher in clinical trials than in non-research settings.[155] Although there is no consensus on the acceptable minimum adherence level in clinical trials, low adherence can have a substantial effect on statistical power and interpretation of trial results.[156-158] Since fewer participants are receiving the full intervention as intended, non-adherence can reduce the contrast between study groups—leading to decreased study power and increased costs associated with recruiting larger sample sizes for evaluating superiority, or leading to potentially inappropriate conclusions of non-inferiority or equivalence. There is also the possibility of underestimating any efficacy and harms of the study intervention.

Furthermore, if adherence is a marker for general healthy behaviour associated with better prognosis, then different rates of non-adherence between study groups can lead to a biased estimate of an intervention's effect. In support of this "healthy adherer" effect, non-adherers to placebo in clinical studies have been found to have poorer clinical outcomes than adherers.[159]

To help avoid these potential detrimental effects of non-adherence, many trials implement procedures and strategies for monitoring and improving adherence,[67 156-158] and any such plans should be described in the protocol.[160] Among applicable drug trials published in 1997-99, 47% reported monitoring the level of adherence.[161] Although each of the many types of monitoring methods has its limitations,[157 158] adherence data can help to inform the statistical analysis (Item 20c), trial interpretation, and choice of appropriate adherence strategies to implement in the trial as it progresses or in future trials and clinical practice.

A variety of adherence strategies exist,[156-158] and their use can be tailored to the specific type of trial design, intervention, and participant population. It may be desirable to select strategies that can be easily implemented in clinical practice, so that the level of adherence in the real world setting is comparable to that observed in the trial.[158]

Interventions—concomitant care

Item 11d: Relevant concomitant care and interventions that are permitted or prohibited during the trial

Explanation
In a controlled trial, a key goal is to have comparable study groups that differ only by the intervention being evaluated, so that any difference in outcomes can be attributed to effects of the study intervention. Cointervention bias can arise when the study groups receive different concomitant care or interventions (in addition to the assigned trial interventions) that may affect trial outcomes.[162] To promote comparability of study groups, the protocol should list the relevant concomitant care and interventions that are allowed (including rescue interventions), as well as any that are prohibited.

Example

"2. Rescue Medication

For weeks 0-3, topical mometasone furoate 0.1% cream or ointment (30 g/week) will be permitted with participants preferably using ointment. Participants will be instructed to apply the topical mometasone furoate to blisters/lesions as required (not to areas of unaffected skin). If the participant is allergic to mometasone furoate or the hospital pharmacy does not stock it, then an alternative topical steroid may be prescribed but this must be in the potent class. In addition, participants will be advised that they can apply a light moisturiser to blisters/lesions at any time during the study.

For weeks 3-6, use of mometasone furoate (or other topical corticosteroids) is strongly discouraged to prevent potential systemic effects. Accidental use of mometasone furoate or other potent topical steroid during this period will be classified as a protocol deviation.

After week 6, potent topical corticosteroids (up to 30 g/week) may be used to treat symptoms and localised disease if they would have normally been used as part of normal clinical care by the physician in charge of that patient. This must be recorded on the trial treatment log.

However, those patients who are on a dose reducing regime for oral steroids, 30 g/week of a "potent" topical steroid will be allowed.

3. Prohibited Concomitant Medications

The administration of live virus vaccines is not permitted for all participants during weeks 0-6 as the investigator is blinded to treatment allocation, and must therefore warn all participants to refrain for [sic] having a live virus vaccine. However, after week 6, once the investigator knows which medication the participant is on, only those taking prednisolone will not be allowed live virus vaccines. Participants should continue to take medications for other conditions as normal. However, if it is anticipated that the participant will need a live virus vaccine during the intervention phase, they will be ineligible for entry into the study . . ."[50]

Outcomes

Item 12: Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended

Example

"1. Primary Outcome Measures

- Difference between the two treatment arms in the proportion of participants classed as treatment success at 6 weeks. Treatment success is defined as 3 or less significant blisters present on examination at 6 weeks. Significant blisters are defined as intact blisters containing fluid which are at least 5 mm in diameter. However, if the participant has popped a blister, or the blister is at a site that makes it susceptible to bursting such as the sole of the foot, it can be considered part of the blister count, providing there is a flexible (but not dry) roof present over a moist base. Mucosal blisters will be excluded from the count.

A survey of the UK DCTN [Dermatology Clinical Trials Network] membership showed that a point estimate of 25% inferiority in effectiveness would be acceptable assuming a gain in the safety profile of at least 10%.

- This measure of success was selected as it was considered to be more clinically relevant than a continuous measure of blister count. It would be less clinically relevant to perform an absolute blister count and report a percentage reduction. Instead, to state that treatment is considered a success if remission is achieved (ie the presence of three or less blisters on physical examination at 6 weeks) more closely reflects clinical practice. In addition, it is far less burdensome on investigators than including a full blister count, which would mean counting in the region of 50-60 blisters in many cases. This outcome measure will be performed as a single blind assessment.
- Difference between the two treatment arms in the proportion of participants reporting grade 3, 4 and 5 (mortality) adverse events which are possibly, probably or definitely related to BP [bullous pemphigoid] medication in the 52 weeks following randomisation. A modified version of The Common Terminology Criteria for Adverse Events (CTCAE v3.0) will be used to grade adverse events. At each study visit, participants will be questioned about adverse events they have experienced since the last study visit (using a standard list of known side effects of the two study drugs).

2. Secondary Outcome Measures

For the secondary and tertiary endpoints a participant will be classed as a treatment success if they have 3 or less significant blisters present on examination and have not had their treatment modified (changed or dose increased) on account of a poor response.

- Difference in the proportion of participants who are classed as a treatment success at 6 weeks.
- Difference in the proportion of participants in each treatment arm who are classed as treatment success at 6 weeks and are alive at 52 weeks. This measure will provide a good overall comparison of the two treatment arms."[50]

Explanation

The trial outcomes are fundamental to study design and interpretation of results. For a given intervention, an outcome can generally reflect efficacy (beneficial effect) or harm (adverse effect). The outcomes of main interest are designated as primary outcomes, which usually appear in the objectives (Item 7) and sample size calculation (Item 14). The remaining outcomes constitute secondary or other outcomes.

For each outcome, the trial protocol should define four components: the specific measurement variable, which corresponds to the data collected directly from trial participants (eg, Beck Depression Inventory score, all cause mortality); the participant-level analysis metric, which corresponds to the format of the outcome data that will be used from each trial participant for analysis (eg, change from baseline, final value, time to event); the method of aggregation, which refers to the summary measure format for each study group (eg, mean, proportion with score > 2); and the specific measurement time point of interest for analysis.[163]

It is also important to explain the rationale for the choice of trial outcomes. An ideal outcome is valid, reproducible, relevant to the target population (eg, patients), and responsive to changes in the health condition being studied.[67] The use of a continuous versus dichotomous method of aggregation can affect study power and estimates of treatment effect,[164 165] and subjective outcomes are more prone to bias from inadequate blinding (ascertainment bias) and allocation concealment (selection bias) than objective outcomes.[166 167] Although composite outcomes increase event rates and statistical power, their relevance and interpretation can be unclear if the individual component outcomes vary greatly in event rates, importance to patients, or amount of missing data.[168-171]

The number of primary outcomes should be as small as possible. Although up to 38% of trials define multiple primary outcomes,[4 35 163] this practice can introduce problems with multiplicity, selective reporting, and interpretation when there are inconsistent results across outcomes. Problems also arise when trial protocols do not designate any primary outcomes, as seen in half (28/59) of protocols for a sample of trials published from 2002-2008,[12] and in 25% of randomised trial protocols that received ethics approval in Denmark in 1994-95.[4] Furthermore, major discrepancies in the primary outcomes designated in protocols/registries/regulatory submissions versus final trial publications are common; favour the reporting of statistically significant primary outcomes over non-significant ones; and are often not acknowledged in final publications.[172-176] Such bias can only be identified and deterred if trial outcomes are clearly defined beforehand in the protocol and if protocol information is made public.[177]

Where possible, the development and adoption of a common set of key trial outcomes within a specialty can help to deter selective reporting of outcomes and to facilitate comparisons and pooling of results across trials in a meta-analysis.[178-180] The COMET (Core Outcome Measures in Effectiveness Trials) Initiative aims to facilitate the development and application of such standardised sets of core outcomes for clinical trials of specific conditions (www.comet-initiative.org). Trial investigators are encouraged to ascertain whether there is a core outcome set relevant to their trial and, if so, to include those outcomes in their trial. Existence of a common set of outcomes does not preclude inclusion of additional relevant outcomes for a given trial.

Participant timeline

Item 13: Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see fig 1)

Examples

"The main outcomes of interest are the drug and sex-related HIV and HCV [hepatitis C virus] risk behaviors . . . Clients will be assessed using the full battery of instruments from the Common Assessment Battery (CAB), along with the Self-Efficacy and Stages of Change questionnaires and a Urine Drug Screen after consenting . . . questionnaires will take place for all participants 14-30 days after randomization during which they will be given the Stages of Change and Self-Efficacy questionnaires, the Timeline Follow-Back, and a UA [urine analysis]. Follow-up interviews, using the full battery (CAB and questionnaires), will be collected at 2 months (56 days), 4 months (112 days) and 6 months (168 days) after the randomization date. A 14 day window, defined as 7 days before and 7 days after the due date, will be available to complete the 2 and 4 month follow-up interviews and a 28 day window, defined as 7 days before and 21 days after the due date, will be available to complete the 6 month follow up interview . . .

7.1.1 Common Assessment Battery (CAB)

A Demographic Questionnaire . . .

The Composite International Diagnostic Interview Version 2.1 . . .

The Addiction Severity Index-Lite (ASI-Lite) . . .

The Risk Behavior Survey (RBS), . . .

7.1.2 Additional Interviews/Questionnaires

To assess drug use, urinalysis for morphine, cocaine, amphetamine, and methamphetamine will be performed at the 2-Week Interim Visit, and the 2-, 4-, and 6-month Follow-up visits . . .

Stage of change for quitting drug use will be measured using a modification of the Motivation Scales [table 3] . . ."[181]

"The trial consists of a 12-week intervention treatment phase with a 40-week follow-up phase. The total trial period will be 12 months. As shown . . . measurements will be undertaken at four time-points in each group: at baseline, directly after completing the 12-week internet program, and at six and 12-month follow-up [see fig 2]."[182]

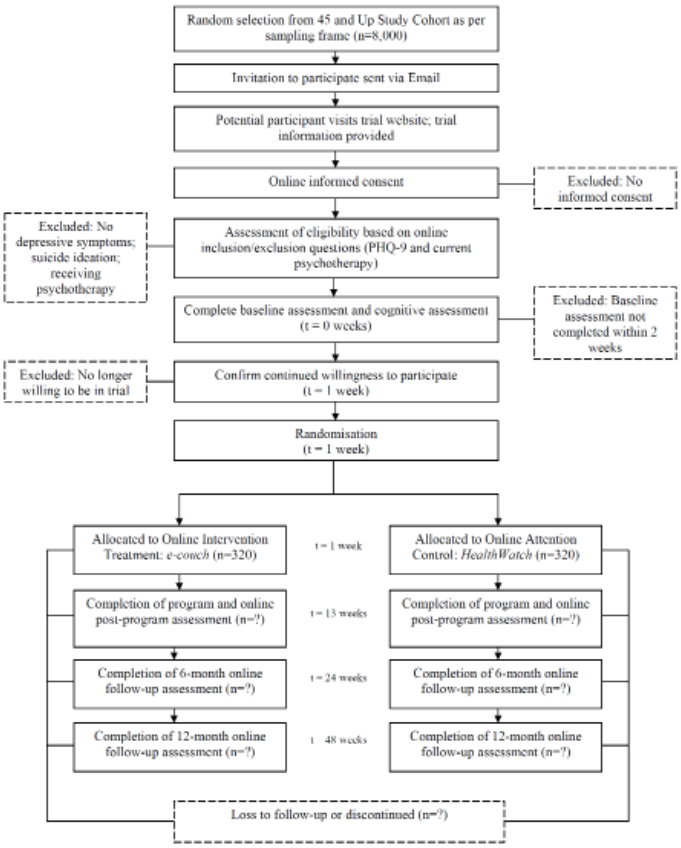


Fig 2: Flow of participants[182]

Explanation

A clear and concise timeline of the study visits, enrolment process, interventions, and assessments performed on participants can help to guide trial conduct and enable external review of participant burden and feasibility. These factors can also affect the decision of potential investigators and participants to join the trial (Item 15).[91]

A schematic diagram is highly recommended to efficiently present the overall schedule and time commitment for trial participants in each study group. Though various presentation formats exist, key information to convey includes the timing of each visit, starting from initial eligibility screening through to study close-out; time periods during which trial interventions will be administered; and the procedures and assessments performed at each visit (with reference to specific data collection forms, if relevant) (fig 1).

Sample size

Item 14: Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations

Explanation

The planned number of trial participants is a key aspect of study design, budgeting, and feasibility that is usually determined using a formal sample size calculation. If the planned sample size is not derived statistically, then this should be explicitly stated along with a rationale for the intended sample size (eg, exploratory nature of pilot studies; pragmatic considerations for trials in rare diseases).[17 184]

For trials that involve a formal sample size calculation, the guiding principle is that the planned sample size should be large enough to have a high probability (power) of detecting a true effect of a given magnitude, should it exist. Sample size calculations are generally based on one primary outcome; however, it may also be worthwhile to plan for adequate study power or report the power that will be available (given the proposed sample size) for other important outcomes or analyses because trials are often underpowered to detect harms or subgroup effects.[185 186]

Among randomised trial protocols that describe a sample size calculation, 4-40% do not state all components of the calculation.[6 11] The protocol should generally include the following: the outcome (Item 12); the values assumed for the outcome in each study group (eg, proportion with event, or mean and standard deviation) (table 4); the statistical test (Item 20a); alpha (type 1 error) level; power; and the calculated sample size per group—both assuming no loss of data and, if relevant, after any inflation for anticipated missing data (Item 20c). Trial investigators are also encouraged

Table 4 | Outcome values to report in sample size calculation

	Type of summary outcome		
Element	Binary	Continuous	Time to event
Assumed result for each study group	Proportion (%) with event	Mean and standard deviation	Proportion (%) with event at a given time point
Effect measure	Relative risk, odds ratio	Difference in means	Hazard ratio

Note: Although the sample size calculation uses the expected outcome value for each group, the corresponding contrast between groups (estimated effect) should also be reported.

Table 3 | HIV/HCV risk reduction protocol schedule of forms and procedures (adapted from original table[181])

Activity/ assessment	CRF (Yes/No)	Staff member	Approximate time to complete (min)	-1 Prestudy screening/ consent	0 Prestudy baseline/ randomisation	T1 Study visit 1	T2 Study/interim visit 2 and/or 2 week interim	F1 Follow-up 2 months	F2 Follow-up 4-months	F3 Follow-up 6-months
Prescreening consent	No	Study coordinator	5	X						
Screening log	No	Study coordinator	5	X						
Consent form/quiz	No	Study coordinator	45	X						
Inclusion/exclusion form	Yes	Study coordinator	N/A	X						
Urine screen	Yes	Study coordinator	10		X		X	X	X	X
Locator form	No	Interviewer	10		X		Update X	Update X	Update X	
Demographics questionnaire	Yes	Interviewer	10		X					
Addiction severity index (ASI) lite	Yes	Interviewer	45		X			X	X	X
Composite international diagnostic interview)	Yes	Interviewer	45		X					
HIV risk behaviour survey	Yes	Interviewer	15		X			X	X	X
Timeline follow back	Yes	Interviewer	5				X	X	X	X
Self efficacy	Yes	Interviewer	5		X		X	X	X	X
Stage of change	Yes	Interviewer	5		X		X	X	X	X
Randomisation	Yes	Study coordinator	15		X					
Voluntary blood sample Counselling and education intervention (treatment group)	Yes	Study phlebotomist	15			X				X
All groups, optional blood sample at study close	Yes	Study phlebotomist	15							X
Termination form	Yes	Study coordinator	N/A							X
Serious adverse event form	Yes	Study coordinator	N/A	As needed throughout protocol						
Progress notes	No	All team members	N/A	X	X	X	X	X	X	X
Communication log	No	All team members	N/A	Every phone or in-person contact outside of a regular visit						

Examples

"The sample size was calculated on the basis of the primary hypothesis. In the exploratory study,[reference] those referred to PEPS [psychoeducation with problem solving] had a greater improvement in social functioning at 6 month follow-up equivalent to 1.05 points on the SFQ [Social Functioning Questionnaire]. However, a number of people received PEPS who were not included in the trial (eg, the wait-list control) and, for this larger sample (N=93), the mean pre-post- treatment difference was 1.79 (pre-treatment mean=13.85, SD=4.21; post-treatment mean=12.06, SD=4.21). (Note: a lower SFQ score is more desirable). This difference of almost 2 points accords with other evidence that this is a clinically significant and important difference.[reference] A reduction of 2 points or more on the SFQ at 1 year follow-up in an RCT of cognitive behaviour therapy in health anxiety was associated with a halving of secondary care appointments (1.24.vs 0.65), a clinically significant reduction in the Hospital Anxiety and Depression Scale (HADS[reference]) Anxiety score of 2.5 (9.9 vs 7.45) and a reduction in health anxiety (the main outcome) of 5.6 points (17.8 vs 12.2) (11 is a normal population score and 18 is pathological).[reference] These findings suggest that improvements in social functioning may accrue over 1 year, hence we expect to find a greater magnitude of response at the 72 week follow-up than we did in the exploratory trial. Therefore, we have powered this trial to be able to detect a difference in SFQ score of 2 points. SFQ standard deviations vary between treatment, control, and the wait-list samples, ranging from 3.78 to 4.53. We have based our sample size estimate on the most conservative (ie, largest) SD [standard deviation]. To detect a mean difference in SFQ score of 2 point (SD = 4.53) at 72 weeks with a two-sided significance level of 1% and power of 80% with equal allocation to two arms would require 120 patients in each arm of the trial. To allow for 30% drop out, 170 will be recruited per arm, ie, 340 in total."[183]

"Superficial and deep incisional surgical site infection rates for patients in the PDS II® [polydioxanone suture] group are estimated to occur at a rate of 0.12.[reference] The trials by [reference] have shown a reduction of SSI [surgical site infections] of more than 50% (from 10.8% to 4.9% and from 9.2% to 4.3% respectively). Therefore, we estimate a rate of 0.06 for PDS Plus® [triclosan-coated continuous polydioxanone suture].

For a fixed sample size design, the sample size required to achieve a power of 1-β=0.80 for the one-sided chi-square test at level α=0.025 under these assumptions amounts to 2×356=712 (nQuery Advisor®, version 7.0). It can be expected that including covariates of prognostic importance in the logistic regression model as defined for the confirmatory analysis will increase the power as compared to the chi-square test. As the individual results for the primary endpoint are available within 30 days after surgery, the drop-out rate is expected to be small. Nevertheless, a potential dilution of the treatment effect due to drop-outs is taken into account (eg no photographs available, loss to follow up); it is assumed that this can be compensated by additional 5% of patients to be randomized, and therefore the total sample size required for a fixed sample size design amounts to n=712+38=750 patients.

…

An adaptive interim analysis [reference] will be performed after availability of the results for the primary endpoint for a total of 375 randomized patients (ie, 50% of the number of patients required in a fixed sample size design). The following type I error rates and decision boundaries for the interim and the final analysis are specified:

- Overall one-sided type I error rate: 0.025
- Boundary for the one-sided p-value of the first stage for accepting the null-hypothesis within the interim analysis: α_0=0.5
- One-sided local type I error rate for testing the null-hypothesis within the interim analysis: α_1=0.0102
- Boundary for the product of the one-sided p-values of both stages for the rejection of the null-hypothesis in the final analysis: cα=0.0038

If the trial will be continued with a second stage after the interim analysis (this is possible if for the one-sided p-value p_1 of the interim analysis $p_1 \in]0.0102, 0.5[$ [ie $0.5 \geq P_1 \geq 0.0102$] holds true, the results of the interim analysis can be taken into account for a recalculation of the required sample size. If the sample size recalculation leads to the conclusion that more than 1200 patients are required, the study is stopped, because the related treatment group difference is judged to be of minor clinical importance.

…

The actually achieved sample size is then not fixed but random, and a variety of scenarios can be considered. If the sample size is calculated under the same assumptions with respect to the SSI rates for the two groups, applying the same the overall significance level of α=0.025 (one-sided) but employing additionally the defined stopping boundaries and recalculating the sample size for the second stage at a conditional power of 80% on the basis of the SSI rates observed in the interim analysis results in an average total sample size of n=766 patients; the overall power of the study is then 90% (ADDPLAN®, version 5.0)."[100]

to provide a rationale or reference for the outcome values assumed for each study group.[187] The values of certain prespecified variables tend to be inappropriately inflated (eg, clinically important treatment effect size)[188 189] or underestimated (eg, standard deviation for continuous outcomes),[190] leading to trials having less power in the end than what was originally calculated. Finally, when uncertainty of a sample size estimate is acknowledged, methods exist for re-estimating sample size.[191] The intended use of such an adaptive design approach should be stated in the protocol.

For designs and frameworks other than parallel group superiority trials, additional elements are required in the sample size calculation. For example, an estimate of the standard deviation of within-person changes from baseline should be included for crossover trials[192]; the intracluster correlation coefficient for cluster randomised trials[193]; and the equivalence or non-inferiority margin for equivalence or non-inferiority trials respectively.[108 194] Such elements are often not described in final trial reports,[110 195-198] and it is unclear how often they are specified in the protocol.

Complete description of sample size calculations in the protocol enables an assessment of whether the trial will be adequately powered to detect a clinically important difference.[189 199-206] It also promotes transparency and discourages inappropriate post hoc revision that is intended to support a favourable interpretation of results or portray consistency between planned and achieved sample sizes.[6 207]

Recruitment

Item 15: Strategies for achieving adequate participant enrolment to reach target sample size

Explanation

The main goal of recruitment is to meet the target sample size (Item 14). However, recruitment difficulties are commonly encountered in clinical trials.[209-213] For example, reviews of government funded trials in the US and UK found that two thirds did not reach their recruitment targets.[214 215] Low enrolment will reduce statistical power and can lead to early trial stoppage or to extensions with delayed results and greater costs.

Strategies to promote adequate enrolment are thus important to consider during trial planning. Recruitment strategies can vary depending on the trial topic, context, and site. Different recruitment methods can substantially affect the number and type of trial participants recruited[128 209 216-220] and can incur different costs.[221-223] Design issues such as the number and stringency of eligibility criteria will also directly affect the number of eligible trial participants.

Protocol descriptions of where participants will be recruited (eg, primary care clinic, community), by whom (eg, surgeon), when (eg, time after diagnosis), and how (eg, advertisements, review of health records) can be helpful for assessing the feasibility of achieving the target sample size and the applicability of the trial results in practice. Other relevant information to explicitly provide in the protocol includes expected recruitment rates, duration of the recruitment period, plans to monitor recruitment during the trial, and any financial or non-financial incentives provided to trial investigators or participants for enrolment (Item 4). If strategies differ by site in multicentre trials, these should be detailed to the extent possible.

Example

"Each center will screen subjects to achieve screening percentages of 50% women and 33% minority; screening will continue until the target population is achieved (12 subjects/site). We recognize that, because of exclusion by genotype and genotypic variation among diverse populations,[reference], the enrolled cohort may not reflect the screened population. The enrollment period will extend over 12 months.

Recruitment Strategy

Each clinical center involved in the ACRN [Asthma Clinical Research Network] was chosen based on documentation for patient availability, among other things. It is, however, worthy to note the specific plans of each center.

. . . The Asthma Clinical Research Center at the Brigham & Women's Hospital utilizes three primary resources for identifying and recruiting potential subjects as described below.

1. Research Patient Database

The Asthma Clinical Research Center at the Brigham and Women's Hospital has a database of over 1,500 asthmatics . . .

2. Asthma Patient Lists . . .

3. Advertisements . . .

. . . the Madison ACRN site has utilized some additional approaches to target minority recruitment. We have utilized a marketing expert to coordinate and oversee our overall efforts in recruiting and retaining minorities. . . . As a result of his efforts, we have advertised widely in newspapers and other publications that target ethnic minorities, established contacts with various ethnic community, university, church, and business groups, and conducted community-based asthma programs . . . For example, student groups such as AHANA (a pre-health careers organization focusing on minority concerns) will be contacted. . . . In addition, we will utilize published examples of successful retention strategies such as frequent payment of subject honoraria as study landmarks are achieved and study participant group social events. Study visits will be carefully planned and scheduled to avoid exam-time and university calendar breaks . . .

The Harlem Hospital Center Emergency Department (ED) sees an average of eight adult patients per day for asthma. Through the REACH (Reducing Emergency Asthma Care in Harlem) project, we have . . . successfully recruited and interviewed 380 patients from the ED . . .

Responses to inquiries about participation in research studies are answered by a dedicated phone line that is manned during business hours and answered by voicemail at all other times. A research assistant responds to each inquiry immediately, using a screening instrument . . .

Patients are recruited for clinical trials at the Jefferson Center through two primary mechanisms: (1) local advertising; and (2) identification in the asthma patient registry (database). Local advertising takes advantage of the printed as well as the audio-visual media. Printed media include . . . All advertising in the printed and audio-visual media has prior approval of the Institutional Review Board.

The Jefferson patient registry (database) has been maintained since 1992 and currently contains 3,100 patients . . . It is estimated that 300-400 new asthmatic patients are seen each year, while a smaller number become inactive due to relocation, change of health care provider, etc. Once identified in the database, patients potentially eligible for a specific study are contacted by the nurse coordinator who explains the study and ascertains the patient's interest. If interested, the patient is seen in the clinical research laboratories where more detailed evaluations are made . . .

Each subject will receive financial compensation within FDA [Food and Drug Administration] guidelines for participation in an amount determined by the local center. For subjects who drop out, payments will be pro-rated for the length of time they stayed in the study, but payment will not be made until the study would have been completed had the subject not dropped out."[208]

Section 3b: Methods—assignment of interventions (for controlled trials)

Allocation—sequence generation

Item 16a: Method of generating the allocation sequence (eg, computer-generated random numbers) and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions

Example

"Participants will be randomly assigned to either control or experimental group with a 1:1 allocation as per a computer generated randomisation schedule stratified by site and the baseline score of the Action Arm Research Test (ARAT; <=21 versus >21) using permuted blocks of random sizes. The block sizes will not be disclosed, to ensure concealment."[224]

Explanation

Participants in a randomised trial should be assigned to study groups using a random (chance) process characterised by unpredictability of assignments. Randomisation decreases selection bias in allocation; helps to facilitate blinding/masking after allocation; and enables the use of probability theory to test whether any difference in outcome between intervention groups reflects chance.[17 225-227] Use of terms such as "randomisation" without further elaboration is not sufficient to describe the allocation process, as these terms have been used inappropriately to describe non-random, deterministic allocation methods such as alternation or allocation by date of birth.[121] In general, these non-random allocation methods introduce selection bias and biased estimates of an intervention's effect size,[17 167 228 229] mainly due to the lack of allocation concealment (Item 16b). If non-random allocation is planned, then the specific method and rationale should be stated. Box 1 outlines the key elements of the random sequence that should be detailed in the protocol. Three quarters of randomised trial protocols approved by a research ethics committee in Denmark (1994-95) or conducted by a US cooperative cancer research group (1968-2006) did not describe the method of sequence generation.[2 11]

Box 1 | Key elements of random sequence to specify in trial protocols

- Method of sequence generation (eg, random number table or computerised random number generator)
- Allocation ratio (Item 8) (eg, whether participants are allocated with equal or unequal probabilities to interventions)
- Type of randomisation (box 2): simple versus restricted; fixed versus adaptive (eg, minimisation); and, where relevant, the reasons for such choices
- If applicable, the factors (eg, recruitment site, sex, disease stage) to be used for stratification (box 2), including categories and relevant cut-off boundaries

Box 2 | Randomisation and minimisation (adapted from CONSORT 2010 Explanation and Elaboration)[17 230 231]

Simple randomisation

Randomisation based solely on a single, constant allocation ratio is known as simple randomisation. Simple randomisation with a 1:1 allocation ratio is analogous to a coin toss, although tossing a coin is not recommended for sequence generation. No other allocation approach, regardless of its real or supposed sophistication, surpasses the bias prevention and unpredictability of simple randomisation.[231]

Restricted randomisation

Any randomised approach that is not simple randomisation is restricted. Blocked randomisation is the most common form. Other forms, used much less frequently, are methods such as replacement randomisation, biased coin, and urn randomisation.[231]

Blocked randomisation

Blocked randomisation (also called permuted block randomisation) assures that study groups of approximately the same size will be generated when an allocation ratio of 1:1 is used. Blocking can also ensure close balance of the numbers in each group at any time during the trial. After every block of eight participants, for example, four would have been allocated to each trial group.[232] Improved balance comes at the cost of reducing the unpredictability of the sequence. Although the order of interventions varies randomly within each block, a person running the trial could deduce some of the next treatment allocations if they discovered the block size.[233] Blinding the interventions, using larger block sizes, and randomly varying the block size will help to avoid this problem.

Biased coin and urn randomisation

Biased coin designs attain the similar objective as blocked designs without forcing strict equality. They therefore preserve much of the unpredictability associated with simple randomisation. Biased-coin designs alter the allocation ratio during the course of the trial to rectify imbalances that might be occurring.[231] Adaptive biased-coin designs, such as the urn design, vary allocation ratios based on the magnitude of the imbalance. However, these approaches are used infrequently.

Stratified randomisation

Stratification is used to ensure good balance of participant characteristics in each group. Without stratification, study groups may not be well matched for baseline characteristics, such as age and stage of disease, especially in small trials. Such imbalances can be avoided without sacrificing the advantages of randomisation. Stratified randomisation is achieved by performing a separate randomisation procedure within each of two or more strata of participants (eg, categories of age or baseline disease severity), ensuring that the numbers of participants receiving each intervention are closely balanced within each stratum. Stratification requires some form of restriction (eg, blocking within strata) in order to be effective. The number of strata should be limited to avoid over-stratification.[234] Stratification by centre is common in multicentre trials.

Minimisation

Minimisation assures similar distribution of selected participant factors between study groups.[230 235] Randomisation lists are not set up in advance. The first participant is truly randomly allocated; for each subsequent participant, the treatment allocation that minimises the imbalance on the selected factors between groups at that time is identified. That allocation may then be used, or a choice may be made at random with a heavy weighting in favour of the intervention that would minimise imbalance (for example, with a probability of 0.8). The use of a random component is generally preferable.[236] Minimisation has the advantage of making small groups closely similar in terms of participant characteristics at all stages of the trial.

Minimisation offers the only acceptable alternative to randomisation, and some have argued that it is superior.[237] On the other hand, minimisation lacks the theoretical basis for eliminating bias on all known and unknown factors. Nevertheless, in general, trials that use minimisation are considered methodologically equivalent to randomised trials, even when a random element is not incorporated. For SPIRIT, minimisation is considered a restricted randomisation approach without any judgment as to whether it is superior or inferior compared to other restricted randomisation approaches.

Box 3 | Need for a separate document to describe restricted randomisation

If some type of restricted randomisation approach is to be used, in particular blocked randomisation or minimisation, then the knowledge of the specific details could lead to bias.[238 239] For example, if the trial protocol for a two arm, parallel group trial with a 1:1 allocation ratio states that blocked randomisation will be used and the block size will be six, then trial implementers know that the intervention assignments will balance every six participants. Thus, if intervention assignments become known after assignment, knowing the block size will allow trial implementers to predict when equality of the sample sizes will arise. A sequence can be discerned from the pattern of past assignments and then some future assignments could be accurately predicted. For example, if part of a sequence contained two "As" and three "Bs," trial implementers would know the last assignment in the sequence would be an "A." If the first three assignments in a sequence contained three "As," trial implementers would know the last three assignments in that sequence would be three "Bs." Selection bias could result, regardless of the effectiveness of allocation concealment (Item 16b).

Of course, this is mainly a problem in open label trials, where everyone becomes aware of the intervention after assignment. It can also be a problem in trials where everyone is supposedly blinded (masked), but the blinding is ineffective or the intervention harms provide clues such that treatments can be guessed.

We recommend that trial investigators do not provide full details of a restricted randomisation scheme (including minimisation) in the trial protocol. Knowledge of these details might undermine randomisation by facilitating deciphering of the allocation sequence. Instead, this specific information should be provided in a separate document with restricted access. However, simple randomisation procedures could be reported in detail in the protocol, because simple randomisation is totally unpredictable.

Table 5 | Differences between allocation concealment and blinding (masking) for trials with individual randomisation

	Allocation concealment	Blinding
Definition	Unawareness of the next study group assignment in the allocation sequence	Unawareness of the study group to which trial participants have already been assigned
Purpose	Prevent selection bias by facilitating enrolment of comparable participants in each study group	Prevent ascertainment, performance, and attrition biases by facilitating comparable concomitant care (aside from trial interventions) and evaluation of participants in each study group
Timing of implementation	Before study group assignment	Upon study group assignment and beyond
Who is kept unaware	Trial participants and individuals enrolling them	One or more of the following: trial participants, investigators, care providers, outcome assessors. Other groups: endpoint adjudication committee, data handlers, data analysts
Always possible to implement?	Yes	No

Box 2 defines the various types of randomisation, including minimisation. When restricted randomisation is used, certain details should not appear in the protocol in order to reduce predictability of the random sequence (box 3). The details should instead be described in a separate document that is unavailable to trial implementers. For blocked randomisation, this information would include details on how the blocks will be generated (eg, permuted blocks by a computer random number generator), the block size(s), and whether the block size will be fixed or randomly varied. Specific block size was provided in 14/102 (14%) randomised trial protocols approved by a Danish research ethics committee in 1994-95, potentially compromising allocation concealment.[2] For trials using minimisation, it is also important to state the details in a separate document, including whether random elements will be used.

Allocation—concealment mechanism

Item 16b: Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned

Example

"Participants will be randomised using TENALEA, which is an online, central randomisation service . . . Allocation concealment will be ensured, as the service will not release the randomisation code until the patient has been recruited into the trial, which takes place after all baseline measurements have been completed."[240]

Explanation

Successful randomisation in practice depends on two interrelated aspects: 1) generation of an unpredictable allocation sequence (Item 16a) and 2) concealment of that sequence until assignment irreversibly occurs.[233 241] The allocation concealment mechanism aims to prevent participants and recruiters from knowing the study group to which the next participant will be assigned. Allocation concealment helps to ensure that a participant's decision to provide informed consent, or a recruiter's decision to enrol a participant, is not influenced by knowledge of the group to which they will be allocated if they join the trial.[242] Allocation concealment should not be confused with blinding (masking) (Item 17) (table 5).[243]

Without adequate allocation concealment, even random, unpredictable assignment sequences can be subverted.[233 241] For example, a common practice is to enclose assignments in sequentially numbered, sealed envelopes. However, if the envelopes are not opaque and contents are visible when held up to a light source, or if the envelopes can be unsealed and resealed, then this method of allocation concealment can be corrupted.

Protocols should describe the planned allocation concealment mechanism in sufficient detail to enable assessment of its adequacy. In one study of randomised trial protocols in Denmark, over half did not adequately describe allocation concealment methods.[2] In contrast, central randomisation was stated as the allocation concealment method in all phase III trial protocols initiated in 1968-2003 by a cooperative cancer research group that used extensive protocol review processes.[11] Like sequence generation, inadequate reporting of allocation concealment in trial publications is common and has been associated with inflated effect size estimates.[167 244 245]

Allocation—implementation

Item 16c: Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions

Explanation

Based on the risk of bias associated with some methods of sequence generation and inadequate allocation concealment, trial investigators should strive for complete separation of the individuals involved in the steps before enrolment (sequence generation process and allocation concealment mechanism) from those involved in the implementation of study group assignments. When this separation is not possible, it is important for the investigators to

Example

"Randomization

All patients who give consent for participation and who fulfil the inclusion criteria will be randomized. Randomisation will be requested by the staff member responsible for recruitment and clinical interviews from CenTrial [Coordination Centre of Clinical Trials].

In return, CenTrial will send an answer form to the study therapist who is not involved in assessing outcome of the study. This form will include a randomisation number. In every centre closed envelopes with printed randomisation numbers on it are available. For every randomisation number the corresponding code for the therapy group of the randomisation list will be found inside the envelopes. The therapist will open the envelope and will find the treatment condition to be conducted in this patient. The therapist then gives the information about treatment allocation to the patient. Staff responsible for recruitment and symptom ratings is not allowed to receive information about the group allocation.

...

The allocation sequence will be generated by the Institute for Medical Biometry (IMB) applying a permuted block design with random blocks stratified by study centre and medication compliance (favourable vs. unfavourable). ... The block size will be concealed until the primary endpoint will be analysed. Throughout the study, the randomisation will be conducted by CenTrial in order to keep the data management and the statistician blind against the study condition as long as the data bank is open. The randomisation list remains with CenTrial for the whole duration of the study. Thus, randomisation will be conducted without any influence of the principal investigators, raters or therapists."[246]

ensure that the assignment schedule is unpredictable and locked away from even the person who generated it. The protocol should specify who will implement the various stages of the randomisation process, how and where the allocation list will be stored, and mechanisms employed to minimise the possibility that those enrolling and assigning participants will obtain access to the list.

Blinding (masking)

Item 17a: Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts) and how

Example

"Assessments regarding clinical recovery will be conducted by an assessor blind to treatment allocation. The assessor will go through a profound assessment training program ... Due to the nature of the intervention neither participants nor staff can be blinded to allocation, but are strongly inculcated not to disclose the allocation status of the participant at the follow up assessments. An employee outside the research team will feed data into the computer in separate datasheets so that the researchers can analyse data without having access to information about the allocation."[247]

Explanation

Blinding or masking (the process of keeping the study group assignment hidden after allocation) is commonly used to reduce the risk of bias in clinical trials with two or more study groups.[166 248] Awareness of the intervention assigned to participants can introduce ascertainment bias in the measurement of outcomes, particularly subjective ones (eg, quality of life)[166 167]; performance bias in the decision to discontinue or modify study interventions (eg, dosing changes) (Item 11b), concomitant interventions, or other aspects of care (Item 11d)[229]; and exclusion/attrition bias in the decision to withdraw from the trial or to exclude a participant from the analysis.[249 250] We have elected to use the term "blinding" but acknowledge that others prefer the term "masking" because "blind" also relates to an ophthalmological condition and health outcome.[251 252]

Many groups can be blinded: trial participants, care providers, data collectors, outcome assessors or committees (Item 5d), data analysts,[253] and manuscript writers. Blinding of data monitoring committees is generally discouraged.[254 255]

When blinding of trial participants and care providers is not possible because of obvious differences between the interventions,[256 257] blinding of the outcome assessors can often still be implemented.[17] It may also be possible to blind participants or trial personnel to the study hypothesis in terms of which intervention is considered active. For example, in a trial evaluating light therapy for depression, participants were informed that the study involved testing two different forms of light therapy, whereas the true hypothesis was that bright blue light was considered potentially effective and that dim red light was considered placebo.[258]

Despite its importance, blinding is often poorly described in trial protocols.[3] The protocol should explicitly state who will be blinded to intervention groups—at a minimum, the blinding status of trial participants, care providers, and outcome assessors. Such a description is much preferred over the use of ambiguous terminology such as "single blind" or "double blind."[259 260] Protocols should also describe the comparability of blinded interventions (Item 11a)[150]—for example, similarities in appearance, use of specific flavours to mask a distinctive taste—and the timing of final unblinding of all trial participants (eg, after the creation of a locked analysis data set).[3]

Furthermore, any strategies to reduce the potential for unblinding should be described in the protocol, such as pretrial testing of blinding procedures.[261] The use of a fixed code (versus a unique code for each participant) to denote each study group assignment (eg, A=Group 1; B=Group 2) can be problematic, as the unblinding of one participant will result in the inadvertent loss of blinding for all trial participants.

Some have suggested that the success of blinding be formally tested by asking key trial persons to guess the study group assignment and comparing these responses to what would be expected by chance.[262] However, it is unclear how best to interpret the results of such tests.[263 264] If done, the planned testing methods should be described in the trial protocol.

Blinding (masking)—emergency unblinding

Item 17b: If blinded, circumstances under which unblinding is permissible and procedure for revealing a participant's allocated intervention during the trial

Explanation

Among 58 blinded Danish trials approved in 1994-95, three quarters of protocols described emergency unblinding procedures.[3] Such procedures to reveal the assigned intervention in certain circumstances are intended to increase the safety of trial participants by informing the clinical management of harms or other relevant conditions that arise. A clear protocol description of the conditions and procedures for emergency unblinding helps to prevent unnecessary unblinding; facilitates implementation by trial personnel when indicated; and enables evaluation of the appropriateness of the planned procedures. In some cases (eg, minor, reversible harms), stopping and then cautiously reintroducing the assigned intervention in the affected participant can avoid both unblinding and further harm.

Example

"To maintain the overall quality and legitimacy of the clinical trial, code breaks should occur only in exceptional circumstances when knowledge of the actual treatment is absolutely essential for further management of the patient. Investigators are encouraged to discuss with the Medical Advisor or PHRI [Population Health Research Institute] physician if he/she believes that unblinding is necessary.

If unblinding is deemed to be necessary, the investigator should use the system for emergency unblinding through the PHRI toll-free help line as the main system or through the local emergency number as the back-up system.

The Investigator is encouraged to maintain the blind as far as possible. The actual allocation must NOT be disclosed to the patient and/or other study personnel including other site personnel, monitors, corporate sponsors or project office staff; nor should there be any written or verbal disclosure of the code in any of the corresponding patient documents.

The Investigator must report all code breaks (with reason) as they occur on the corresponding CRF [case report form] page.

Unblinding should not necessarily be a reason for study drug discontinuation."[265]

Section 3c: Methods—data collection, management, and analysis

Data collection methods

Item 18a: Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol

Explanation

The validity and reliability of trial data depend on the quality of the data collection methods. The processes of acquiring and recording data often benefit from attention to training of study personnel and use of standardised, pilot tested methods. These should be identical for all study groups, unless precluded by the nature of the intervention.

The choice of methods for outcome assessment can affect study conduct and results.[268-273] Substantially different responses can be obtained for certain outcomes (eg, harms) depending on who answers the questions (eg, the participant or investigator) and how the questions are presented (eg, discrete options or open ended).[269 274-276] Also, when compared to paper based data collection, the use of electronic handheld devices and internet websites has the potential to improve protocol adherence, data accuracy, user acceptability, and timeliness of receiving data.[268 270 271 277]

The quality of data also depends on the reliability, validity, and responsiveness of data collection instruments such as questionnaires[278] or laboratory instruments. Instruments with low inter-rater reliability will reduce

Examples

"Primary outcome

Delirium recognition: In accordance with national guidelines [reference], the study will identify delirium by using the RASS [Richmond Agitation-Sedation Scale] and the CAM-ICU [Confusion Assessment Method for the intensive care unit] on all patients who are admitted directly from the emergency room or transferred from other services to the ICU. Such assessment will be performed after 24 hours of ICU admission and twice daily until discharge from the hospital . . . RASS has excellent inter-rater reliability among adult medical and surgical ICU patients and has excellent validity when compared to a visual analogue scale and other selected sedation scales[reference] . . . The CAM-ICU was chosen because of its practical use in the ICU wards, its acceptable psychometric properties, and based on the recommendation of national guidelines[reference] . . . The CAM-ICU diagnosis of delirium was validated against the DSM-III-R [Diagnostic and Statistical Manual of Mental Disorders, Third Edition—Revised] delirium criteria determined by a psychiatrist and found to have a sensitivity of 97% and a specificity of 92%. [reference] The CAM-ICU has been developed, validated and applied into ICU settings and multiple investigators have used the same method to identify patients with delirium.[reference]

Delirium severity: Since the CAM-ICU does not evaluate delirium severity, we selected the Delirium Rating Scale revised-1998 (DRS-R-98)[reference] . . . The DRS-R-98 was designed to evaluate the breadth of delirium symptoms for phenomenological studies in addition to measuring symptom severity with high sensitivity and specificity . . . The DRS-R-98 is a 16-item clinician-rated scale with anchored items descriptions . . . The DRS-R-98 has excellent inter-rater reliability (intra-class correlation 0.97) and internal consistency (Cronbach's alpha 0.94).[reference]

Secondary outcomes

The study will collect demographic and baseline functional information from the patient's legally authorized representative and/or caregivers. Cognitive function status will be obtained by interviewing the patient's legally authorized representative using the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE). IQCODE is a questionnaire that can be completed by a relative or other caregiver to determine whether that person has declined in cognitive functioning. The IQCODE lists 26 everyday situations . . . Each situation is rated by the informant for amount of change over the previous 10 years, using a Likert scale ranging from 1-much improved to 5-much worse. The IQCODE has a sensitivity between 69% to 100% and specificity of 80% to 96% for dementia.[reference]

Utilizing the electronic medical record system (RMRS), we will collect several data points of interest at baseline and throughout the study period . . . We have previously defined hospital-related consequences to include: the number of patients with documented falls, use of physical restraints . . . These will be assessed using the RMRS, direct daily observation, and retrospective review of the electronic medical record. This definition of delirium related hospital complications has been previously used and published.[reference]"[266]

"Training and certification plans

. . . Each center's personnel will be trained centrally in the study requirements, standardized measurement of height, weight, and blood pressure, requirements for laboratory specimen collection including morning urine samples, counseling for adherence and the eliciting of information from study participants in a uniform reproducible manner.

. . . The data to be collected and the procedures to be conducted at each visit will be reviewed in detail. Each of the data collection forms and the nature of the required information will be discussed in detail on an item by item basis. Coordinators will learn how to code medications using the WHODrug software and how to code symptoms using the MedDRA software. Entering data forms, responding to data discrepancy queries and general information about obtaining research quality data will also be covered during the training session.

. . .

13.7. Quality Control of the Core Lab

Data from the Core Lab will be securely transmitted in batches and quality controlled in the same manner as Core Coordinating Center data; ie data will be entered and verified in the database on the Cleveland Clinic Foundation SUN with a subset later selected for additional quality control. Appropriate edit checks will be in place at the key entry (database) level.

The Core Lab is to have an internal quality control system established prior to analyzing any FSGS [focal segmental glomerulosclerosis] samples. This system will be outlined in the Manual of Operations for the Core Lab(s) which is prepared and submitted by the Core Lab to the DCC [data coordinating centre] prior to initiating of the study.

At a minimum this system must include:

1) The inclusion of at least two known quality control samples; the reported measurements of the quality control samples must fall within specified ranges in order to be certified as acceptable.

2) Calibration at FDA approved manufacturers' recommended schedules.

13.8. Quality Control of the Biopsy Committee

The chair of the pathology committee will circulate to all of the study pathologists . . . samples [sic] biopsy specimens for evaluation after criteria to establish diagnosis of FSGS has been agreed. This internal review process will serve to ensure common criteria and assessment of biopsy specimens for confirmation of diagnosis of FSGS."[267]

statistical power,[272] while those with low validity will not accurately measure the intended outcome variable. One study found that only 35% (47/133) of randomised trials in acute stroke used a measure with established reliability or validity.[279] Modified versions of validated measurement tools may no longer be considered validated, and use of unpublished measurement scales can introduce bias and inflate treatment effect sizes.[280]

Standard processes should be implemented by local study personnel to enhance data quality and reduce bias by detecting and reducing the amount of missing or incomplete data, inaccuracies, and excessive variability in measurements.[281-285] Examples include standardised training and testing of outcome assessors to promote consistency; tests of the validity or reliability of study instruments; and duplicate data measurements.

A clear protocol description of the data collection process—including the personnel, methods, instruments, and measures to promote data quality—can facilitate implementation and helps protocol reviewers to assess their appropriateness. Inclusion of data collection forms in the protocol (ie, as appendices) is highly recommended, as the way in which data are obtained can substantially affect the results. If not included in the protocol, then a reference to where the forms can be found should be provided. If performed, pilot testing and assessment of reliability and validity of the forms should also be described.

Data collection methods—retention

Item 18b: Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols

Explanation

Trial investigators must often seek a balance between achieving a sufficiently long follow-up for clinically relevant outcome measurement,[122 288] and a sufficiently short follow-up to decrease attrition and maximise completeness of data collection. Non-retention refers to instances where participants are prematurely "off-study" (ie, consent withdrawn or lost to follow-up) and thus outcome data cannot be obtained from them. The majority of trials will have some degree of non-retention, and the number of these "off-study" participants usually increases with the length of follow-up.[116]

It is desirable to plan ahead for how retention will be promoted in order to prevent missing data and avoid the associated complexities in both the study analysis (Item 20c) and interpretation. Certain methods can improve participant retention,[67 152 289-292] such as financial reimbursement; systematic methods and reminders for contacting patients, scheduling appointments, and monitoring retention; and limiting participant burden related to follow-up visits and procedures (Item 13). A participant who withdraws consent for follow-up assessment of one outcome may be willing to continue with assessments for other outcomes, if given the option.

Non-retention should be distinguished from non-adherence.[293] Non-adherence refers to deviation from intervention protocols (Item 11c) or from the follow-up schedule of assessments (Item 13), but does not mean that the participant is "off-study" and no longer in the trial. Because missing data can be a major threat to trial validity and statistical power, non-adherence should not be an automatic reason for ceasing to collect data from the trial participant prior to study completion. In particular for randomised trials, it is widely recommended that all participants be included in an intention to treat analysis, regardless of adherence (Item 20c).

Examples

"5.2.2 Retention

. . . As with recruitment, retention addresses all levels of participant.

At the parent and student level, study investigators and staff:

- Provide written feedback to all parents of participating students about the results of the "health screenings" . . .
- Maintain interest in the study through materials and mailings . . .
- Send letters to parents and students prior to the final data collection, reminding them of the upcoming data collection and the incentives the students will receive.

At the school level, study investigators and staff:

- Provide periodic communications via newsletters and presentations to inform the school officials/staff, students, and parents about type 2 diabetes, the current status of the study, and plans for the next phase, as well as to acknowledge their support.
- . . .
- Become a presence in the intervention schools to monitor and maintain consistency in implementation, . . . be as flexible as possible with study schedule and proactive in resolving conflicts with schools.
- Provide school administration and faculty with the schedule or grid showing how the intervention fits into the school calendar . . .
- Solicit support from parents, school officials/staff, and teachers . . .
- Provide periodic incentives for school staff and teachers.
- Provide monetary incentives for the schools that increase with each year of the study [table 6]. . ."[286]

"5.4 Infant Evaluations in the Case of Treatment Discontinuation or Study Withdrawal

All randomized infants completing the 18-month evaluation schedule will have fulfilled the infant clinical and laboratory evaluation requirements for the study. . .

All randomized infants who are prematurely discontinued from study drug will be considered *off study drug/on study* and will follow the same schedule of events as those infants who continue study treatment except adherence assessment. All of these infants will be followed through 18 months as scheduled.

Randomized infants prematurely discontinued from the study before the 6-month evaluation will have the following clinical and laboratory evaluations performed, if possible: . . .

- Roche Amplicor HIV-1 DNA PCR [polymerase chain reaction] and cell pellet storage
- Plasma for storage (for NVP [nevirapine] resistance, HIV-1 RNA PCR and NVP concentration)

. . .

Randomized infants prematurely discontinued from the study at any time after the 6-month evaluation will have the following clinical and laboratory evaluations performed, if possible:

. . .

5.5 Participant Retention

Once an infant is enrolled or randomized, the study site will make every reasonable effort to follow the infant for the entire study period . . . It is projected that the rate of loss-to-follow-up on an annual basis will be at most 5% . . . Study site staff are responsible for developing and implementing local standard operating procedures to achieve this level of follow-up.

5.6 Participant Withdrawal

Participants may withdraw from the study for any reason at any time. The investigator also may withdraw participants from the study in order to protect their safety and/or if they are unwilling or unable to comply with required study procedures after consultation with the Protocol Chair, National Institutes of Health (NIH) Medical Officers, Statistical and Data Management Center (SDMC) Protocol Statistician, and Coordinating and Operations Center (CORE) Protocol Specialist.

Participants also may be withdrawn if the study sponsor or government or regulatory authorities terminate the study prior to its planned end date.

Note: Early discontinuation of study product for any reason is not a reason for withdrawal from the study."[287]

Table 6 | Excerpts from table showing compensation provided in study[286]

Who	What	Amount
School		
Intervention school	School program enhancement	$2000 in year 1, $3000 in year 2, $4000 in year 3
	Physical education class equipment required to implement intervention	$15 000 over 3 years
	Food service department to defray costs of nutrition intervention	$3000/year
Control school	School program enhancement	$2000 in year 1, $4000 in year 2, $6000 in year 3
Student		
All	Return consent form (signed or not)	Gift item worth ~ $5
	Participation in health screening data collection measures and forms	$50 baseline (6th grade), $10 interim (7th grade), $60 end of study (8th grade)
Family		
Intervention parents	Focus groups to provide input about family outreach events and activities	$35/year per parent, up to two focus groups per field center, 6-10 participants per focus group

Protocols should describe any retention strategies and define which outcome data will be recorded from protocol non-adherers.[152] Protocols should also detail any plans to record the reasons for non-adherence (eg, discontinuation of intervention due to harms versus lack of efficacy) and non-retention (ie, consent withdrawn; lost to follow-up), as this information can influence the handling of missing data and interpretation of results.[152 294 295]

Data management

Item 19: Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol

Explanation

Careful planning of data management with appropriate personnel can help to prevent flaws that compromise data validity. The protocol should provide a full description of the data entry and coding processes, along with measures to promote their quality, or provide key elements and a reference to where full information can be found. These details are particularly important for the primary outcome data. The protocol should also document data security measures to prevent unauthorised access to or loss of participant data, as well as plans for data storage

Example

"13.9.2. Data Forms and Data Entry

In the FSGS-CT [focal segmental glomerulosclerosis—clinical trial], all data will be entered electronically. This may be done at a Core Coordinating Center or at the participating site where the data originated. Original study forms will be entered and kept on file at the participating site. A subset will be requested later for quality control; when a form is selected, the participating site staff will pull that form, copy it, and sent [sic] the copy to the DCC [data coordinating center] for re-entry.

. . . Participant files are to be stored in numerical order and stored in a secure and accessible place and manner. Participant files will be maintained in storage for a period of 3 years after completion of the study.

13.9.3. Data Transmission and Editing

The data entry screens will resemble the paper forms approved by the steering committee. Data integrity will be enforced through a variety of mechanisms. Referential data rules, valid values, range checks, and consistency checks against data already stored in the database (ie, longitudinal checks) will be supported. The option to chose [sic] a value from a list of valid codes and a description of what each code means will be available where applicable. Checks will be applied at the time of data entry into a specific field and/or before the data is written (committed) to the database. Modifications to data written to the database will be documented through either the data change system or an inquiry system. Data entered into the database will be retrievable for viewing through the data entry applications. The type of activity that an individual user may undertake is regulated by the privileges associated with his/her user identification code and password.

13.9.4. Data Discrepancy Inquiries and Reports to Core Coordinating Centers

Additional errors will be detected by programs designed to detect missing data or specific errors in the data. These errors will be summarized along with detailed descriptions for each specific problem in Data Query Reports, which will be sent to the Data Managers at the Core Coordinating Centers . . .

The Data Manager who receives the inquiry will respond by checking the original forms for inconsistency, checking other sources to determine the correction, modifying the original (paper) form entering a response to the query. Note that it will be necessary for Data Managers to respond to each inquiry received in order to obtain closure on the queried item.

The Core Coordinating Center and participating site personnel will be responsible for making appropriate corrections to the original paper forms whenever any data item is changed . . . Written documentation of changes will be available via electronic logs and audit trails.

. . .

Biopsy and biochemistry reports will be sent via e-mail when data are received from the Core Lab.

. . .

13.9.5. Security and Back-Up of Data

. . . All forms, diskettes and tapes related to study data will be kept in locked cabinets. Access to the study data will be restricted. In addition, Core Coordinating Centers will only have access to their own center's data. A password system will be utilized to control access . . . These passwords will be changed on a regular basis. All reports prepared by the DCC will be prepared such that no individual subject can be identified.

A complete back up of the primary DCC database will be performed twice a month. These tapes will be stored off-site in a climate-controlled facility and will be retained indefinitely. Incremental data back-ups will be performed on a daily basis. These tapes will be retained for at least one week on-site. Back-ups of periodic data analysis files will also be kept. These tapes will be retained at the off-site location until the Study is completed and the database is on file with NIH [National Institutes of Health]. In addition to the system back-ups, additional measures will be taken to back-up and export the database on a regular basis at the database management level. . .

13.9.6. Study status reports

The DCC will send weekly email reports with information on missing data, missing forms, and missing visits. Personnel at the Core Coordinating Center and the Participating Sites should review these reports for accuracy and report any discrepancies to the DCC.

. . .

13.9.8. Description of Hardware at DCC

A SUN Workstation environment is maintained in the department with a SUN SPARCstation 10 model 41 as the server . . . Primary access to the departments [sic] computing facilities will be through the Internet . . . For maximum programming efficiency, the Oracle database management system and the SAS and BMDP statistical analysis systems will be employed for this study. . . . Oracle facilitates sophisticated integrity checks through a variety of mechanisms including stored procedures, stored triggers, and declarative database integrity—for between table verifications. Oracle allows data checks to be programmed once in the database rather than repeating the same checks among many applications . . . Security is enforced through passwords and may be assigned at different levels to groups and individuals."[267]

(including timeframe) during and after the trial. This information facilitates an assessment of adherence to applicable standards and regulations.

Differences in data entry methods can affect the trial in terms of data accuracy,[268] cost, and efficiency.[271] For example, when compared with paper case report forms, electronic data capture can reduce the time required for data entry, query resolution, and database release by combining data entry with data collection (Item 18a).[271 277] When data are collected on paper forms, data entry can be performed locally or at a central site. Local data entry can enable fast correction of missing or inaccurate data, while central data entry facilitates blinding (masking), standardisation, and training of a core group of data entry personnel.

Raw, non-numeric data are usually coded for ease of data storage, review, tabulation, and analysis. It is important to define standard coding practices to reduce errors and observer variation. When data entry and coding are performed by different individuals, it is particularly important that the personnel use unambiguous, standardised terminology and abbreviations to avoid misinterpretation. As with data collection (Item 18a), standard processes are often implemented to improve the accuracy of data entry and coding.[281 284] Common examples include double data entry[296]; verification that the data are in the proper format (eg, integer) or within an expected range of values; and independent source document verification of a random subset of data to identify missing or apparently erroneous values. Though widely performed to detect data entry errors, the time and costs of independent double data entry from paper forms need to be weighed against the magnitude of reduction in error rates compared to single-data entry.[297-299]

Statistical methods

The planned methods of statistical analysis should be fully described in the protocol. If certain aspects of the analysis plan cannot be prespecified (eg, the method of handling missing data is contingent on examining patterns of "missingness" before study unblinding), then the planned approach to making the final methodological choices should be outlined. Some trials have a separate document—commonly called a statistical analysis plan (SAP)—that fully details the planned analyses. Any SAP should be described in the protocol, including its key elements and where it can be found. As with the protocol, the SAP should be dated, amendments noted and dated, and the SAP authors provided.

Statistical methods—outcomes

Item 20a: Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol

Explanation

The protocol should indicate explicitly each intended analysis comparing study groups. An unambiguous, complete, and transparent description of statistical methods facilitates execution, replication, critical appraisal, and the ability to track any changes from the original pre-specified methods.

Example

"The intervention arm (SMS [short message system (text message)]) will be compared against the control (SOC [standard of care]) for all primary analysis. We will use chi-squared test for binary outcomes, and T-test for continuous outcomes. For subgroup analyses, we will use regression methods with appropriate interaction terms (respective subgroup×treatment group). Multivariable analyses will be based on logistic regression . . . for binary outcomes and linear regression for continuous outcomes. We will examine the residual to assess model assumptions and goodness-of-fit. For timed endpoints such as mortality we will use the Kaplan-Meier survival analysis followed by multivariable Cox proportional hazards model for adjusting for baseline variables. We will calculate Relative Risk (RR) and RR Reductions (RRR) with corresponding 95% confidence intervals to compare dichotomous variables, and difference in means will be used for additional analysis of continuous variables. P-values will be reported to four decimal places with p-values less than 0.001 reported as $p < 0.001$. Up-to-date versions of SAS (Cary, NC) and SPSS (Chicago, IL) will be used to conduct analyses. For all tests, we will use 2-sided p-values with alpha≤0.05 level of significance. We will use the Bonferroni method to appropriately adjust the overall level of significance for multiple primary outcomes, and secondary outcomes.

To assess the impact of potential clustering for patients cared by the same clinic, we will use generalized estimating equations [GEE] assuming an exchangeable correlation structure. Table [7] provides a summary of methods of analysis for each variable. Professional academic statisticians (LT, RN) blinded to study groups will conduct all analyses."[47]

Results for the primary outcome can be substantially affected by the choice of analysis methods. When investigators apply more than one analysis strategy for a specific primary outcome, there is potential for inappropriate selective reporting of the most interesting result.[6] The protocol should prespecify the main ("primary") analysis of the primary outcome (Item 12), including the analysis methods to be used for statistical comparisons (Items 20a and 20b); precisely which trial participants will be included (Item 20c); and how missing data will be handled (Item 20c). Additionally, it is helpful to indicate the effect measure (eg, relative risk) and significance level that will be used, as well as the intended use of confidence intervals when presenting results.

The same considerations will often apply equally to prespecified secondary and exploratory outcomes. In some instances, descriptive approaches to evaluating rare outcomes such as adverse events—might be preferred over formal analysis given the lack of power.[300] Adequately powered analyses may require preplanned meta-analyses with results from other studies.

Most trials are affected to some extent by multiplicity issues.[301 302] When multiple statistical comparisons are performed (eg, multiple study groups, outcomes, interim analyses), the risk of false positive (type 1) error is inflated and there is increased potential for selective reporting of favourable comparisons in the final trial report. For trials with more than two study groups, it is important to specify in the protocol which comparisons (of two or more study groups) will be performed and, if relevant, which will be the main comparison of interest. The same principle of specifying the main comparison also applies when there is more than one outcome, including when the same variable

Table 7 | Variables, measures, and methods of analysis (reproduced from original table[47])

Variable/outcome	Hypothesis	Outcome measure	Methods of analysis
1) Primary	Intervention improved outcome from baseline to 6 months		
a) Adherence at 12 months		Percent adherence in previous 30 days >95% [binary]	Chi-squared test
b) Suppression of HIV viral load at 12 months		Viral load ≤400 copies/ml [binary]	Chi-squared test
2) Secondary	improvement occurred	Adherence % (>95%) [binary]	Chi-squared test
Adherence percentage at 12 months			
HIV viral load at 12 months	improvement occurred	Viral load (copies)	T-test
Immune reconstitution (change in CD4 T cellcount from baseline)	improvement occurred	CD4 T-cells/mm^3 (continuous)	T-test
Time to virological failure	improvement occurred	Virological failure after successful suppression	Kaplan-Meier survivalanalysis
Weight gain [lbs] and BMI	improvement occurred	Change in weight (lbs) and BMI [continuous]	T-test
Occurrence of opportunistic infections (OIs)	improvement occurred	Presence of AIDS definingopportunistic infection [binary]	Chi-squared test
Time to reporting of adverse drug events (ADEs)	improvement occurred	Presence of drug-related adverse event [time to event]	Kaplan-Meier survivalanalysis
Deaths (all cause)	improvement occurred	All-cause mortality [binary]	Chi-squared test and Kaplan-Meier survival analysis
SF-12 [short form 12 adapted for regional application in Kiswahili]	improvement occurred	Quality pf [sic] life questionnaire[continuous]	T-test
Satisfaction with care provided	improvement occurred	Questionnaire [continuous]	T-test
Level of disclosure of HIV status	improvement occurred	Disclosed to a family member [binary]	Chi-squared test
Impression of stigma	improvement occurred	Questionnaire [continuous]	T-test
Family dyamics [sic]	improvement occurred	Questionnaire [continuous]	T-test
Employment attendance	improvement occurred	Questionnaire [continuous]	T-test
Household member school attendance	improvement occurred	Questionnaire [continuous]	T-test
Cell phones lost/stolen	improvement occurred	Presence of cellphone [binary]	Poisson regression
Stopped taking HAART [highly active antiretroviral therapy]	improvement occurred	Self-report [binary]	Chi-squared test
Required active tracing for 12 month follow-up	improvement occurred	Field officers [binary]	Chi-squared test
3) Subgroup Analyses:			Regression methods with appropriate interaction term
Urban vs. rural	Distance affectsadherence		
Female vs. male	Sex affects adherence		
Phone ownership (owned vs. shared)	Ownership affectsadherence		
Level of education	Low education affectsadherence		
4) Sensitivity Analyses:	improvement occurred	All outcomes	
a) Per protocol analysis			a) Chi-squared/T-test
b) Adjusting for baseline covariates			b) Multivariable regression
c) clustering among individuals within a clinic			c) GEE

IMPORTANT REMARKS:
- The GEE [generalised estimating equations] [reference] is a technique that allows to specify the correlation structure between patients within a hospital and this approach produces unbiased estimates under the assumption that missing observations will be missing at random. An amended approach of weighted GEE will be employed if missingness is found not to be at random [reference].
- In all analyses results will be expressed as coefficient, standard errors, corresponding 95% and associated p-values.
- Goodness-of-fit will be assessed by examining the residuals for model assumptions and chi-squared test of goodness-of-fit.
- Bonferroni method will be used to adjust the overall level of significance for multiple secondary outcomes.

is measured at several time points (Item 12). Any statistical approaches to account for multiple comparisons and time points should also be described.

Finally, different trial designs dictate the most appropriate analysis plan and any additional relevant information that should be included in the protocol. For example, cluster, factorial, crossover, and within-person randomised trials require specific statistical considerations, such as how clustering will be handled in a cluster randomised trial.

Statistical methods—additional analyses

Item 20b: Methods for any additional analyses (eg, subgroup and adjusted analyses)

Explanation

Subgroup analysis

Subgroup analyses explore whether estimated treatment effects vary significantly between subcategories of trial participants. As these data can help tailor healthcare decisions to individual patients, a modest number of prespecified subgroup analyses can be sensible.

However, subgroup analyses are problematic if they are inappropriately conducted or selectively reported. Subgroup analyses described in protocols or grant applications do not match those reported in subsequent publications for more than two thirds of randomised trials, suggesting that subgroup analyses are often selectively reported or not prespecified.[6 7 305] Post hoc (data driven) analyses have a high risk of spurious findings and are discouraged.[306] Conducting a large number of subgroup comparisons leads to issues of multiplicity, even when all of the comparisons have been pre-specified. Furthermore, when subgroups are based on variables measured after randomisation, the analyses are particularly susceptible to bias.[307]

Preplanned subgroup analyses should be clearly specified in the protocol with respect to the precise baseline variables to be examined, the definition of the subgroup categories (including cut-off boundaries for continuous or ordinal variables), the statistical method to be used, and the hypothesised direction of the subgroup effect based on plausibility.[308 309]

Examples

"We plan to conduct two subgroup analyses, both with strong biological rationale and possible interaction effects. The first will compare hazard ratios of re-operation based upon the degree of soft tissue injury (Gustilo-Anderson Type I/II open fractures vs. Gustilo-Anderson Type IIIA/B open fractures). The second will compare hazard ratios of re-operation between fractures of the upper and lower extremity. We will test if the treatment effects differ with fracture types and extremities by putting their main effect and interaction terms in the Cox regression. For the comparison of pressure, we anticipate that the low/gravity flow will be more effective in the Type IIIA-B open fracture than in the Type I/II open fracture, and be more effective in the upper extremity than the lower extremity. For the comparison of solution, we anticipate that soap will do better in the Type IIIA-B open fracture than in the Type I/II open fracture, and better in the upper extremity than the lower extremity."[303]

"A secondary analysis of the primary endpoint will adjust for those pre-randomization variables which might reasonably be expected to be predictive of favorable outcomes. Generalized linear models will be used to model the proportion of subjects with neurologically intact (MRS ≤ 3 [Modified Rankin Score]) survival to hospital discharge by ITD [impedance threshold device]/sham device group adjusted for site (dummy variables modeling the 11 ROC [Resuscitation Outcomes Consortium] sites), patient sex, patient age (continuous variable), witness status (dummy variables modeling the three categories of unwitnessed arrest, non-EMS [emergency medical services] witnessed arrest, and EMS witnessed arrest), location of arrest (public versus non-public), time or response (continuous variable modeling minutes between call to 911 and arrival of EMS providers on scene), presenting rhythm (dummy variables modeling asystole, PEA [pulseless electrical activity], VT/VF [ventricular tachycardia/fibrillation], or unknown), and treatment assignment in the Analyze Late vs. Analyze Early intervention. The test statistic used to assess any benefit of the ITD relative to the sham device will be computed as the generalized linear model regression coefficient divided by the estimated "robust" standard error based on the Huber- White sandwich estimator[reference] in order to account for within group variability which might depart from the classical assumptions. Statistical inference will be based on one-sided P values and 95% confidence intervals which adjust for the stopping rule used for the primary analysis."[304]

Adjusted analysis

Some trials prespecify adjusted analyses to account for imbalances between study groups (eg, chance imbalance across study groups in small trials), improve power, or account for a known prognostic variable. Adjustment is often recommended for any variables used in the allocation process (eg, in stratified randomisation), on the principle that the analysis strategy should match the design.[310] Most trial protocols and publications do not adequately address issues of adjustment, particularly the description of variables.[6 310]

It is important that trial investigators indicate in the protocol if there is an intention to perform or consider adjusted analyses, explicitly specifying any variables for adjustment and how continuous variables will be handled. When both unadjusted and adjusted analyses are intended, the main analysis should be identified (Item 20a). It may not always be clear, in advance, which variables will be important for adjustment. In such situations, the objective criteria to be used to select variables should be prespecified. As with subgroup analyses, adjustment variables based on post-randomisation data rather than baseline data can introduce bias.[311 312]

Statistical methods—analysis population and missing data

Item 20c: Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)

Explanation

In order to preserve the unique benefit of randomisation as a mechanism to avoid selection bias, an "as randomised" analysis retains participants in the group to which they were originally allocated. To prevent attrition bias, out-

Example

"Nevertheless, we propose to test non-inferiority using two analysis sets; the intention-to-treat set, considering all patients as randomized regardless of whether they received the randomized treatment, and the "per protocol" analysis set. Criteria for determining the "per protocol" group assignment would be established by the Steering Committee and approved by the PSMB [performance and safety monitoring board] before the trial begins. Given our expectation that very few patients will crossover or be lost to follow-up, these analyses should agree very closely. We propose declaring medical management non-inferior to interventional therapy, only if shown to be non-inferior using both the "intention to treat" and "per protocol" analysis sets.

. . .

10.4.7 Imputation Procedure for Missing Data

While the analysis of the primary endpoint (death or stroke) will be based on a log-rank test and, therefore, not affected by patient withdrawals (as they will be censored) provided that dropping out is unrelated to prognosis; other outcomes, such as the Rankin Score at five years post-randomization, could be missing for patients who withdraw from the trial. We will report reasons for withdrawal for each randomization group and compare the reasons qualitatively . . . The effect that any missing data might have on results will be assessed via sensitivity analysis of augmented data sets. Dropouts (essentially, participants who withdraw consent for continued follow-up) will be included in the analysis by modern imputation methods for missing data.

The main feature of the approach is the creation of a set of clinically reasonable imputations for the respective outcome for each dropout. This will be accomplished using a set of repeated imputations created by predictive models based on the majority of participants with complete data. The imputation models will reflect uncertainty in the modeling process and inherent variability in patient outcomes, as reflected in the complete data.

After the imputations are completed, all of the data (complete and imputed) will be combined and the analysis performed for each imputed-and-completed dataset. Rubin's method of multiple (ie, repeated) imputation will be used to estimate treatment effect. We propose to use 15 datasets (an odd number to allow use of one of the datasets to represent the median analytic result).

These methods are preferable to simple mean imputation, or simple "best-worst" or "worst-worst" imputation, because the categorization of patients into clinically meaningful subgroups, and the imputation of their missing data by appropriately different models, accords well with best clinical judgment concerning the likely outcomes of the dropouts, and therefore will enhance the trial's results."[313]

come data obtained from all participants are included in the data analysis, regardless of protocol adherence (Items 11c and 18b).[249 250] These two conditions (ie, all participants, as randomised) define an "intention to treat" analysis, which is widely recommended as the preferred analysis strategy.[17]

Some trialists use other types of data analyses (commonly labelled as "modified intention to treat" or "per protocol") that exclude data from certain participants—such as those who are found to be ineligible after randomisation or who deviate from the intervention or follow-up protocols. This exclusion of data from protocol non-adherers can introduce bias, particularly if the frequency of and the reasons for non-adherence vary between the study groups.[314 315] In some trials, the participants to be included in the analysis will vary by outcome—for example, analysis

of harms (adverse events) is sometimes restricted to participants who received the intervention, so that absence or occurrence of harm is not attributed to a treatment that was never received.

Protocols should explicitly describe which participants will be included in the main analyses (eg, all randomised participants, regardless of protocol adherence) and define the study group in which they will be analysed (eg, as randomised). In one cohort of randomised trials approved in 1994-5, this information was missing in half of the protocols.[6] The ambiguous use of labels such as "intention to treat" or "per protocol" should be avoided unless they are fully defined in the protocol.[6 314] Most analyses labelled as "intention to treat" do not actually adhere to its definition because of missing data or exclusion of participants who do not meet certain post-randomisation criteria (eg, specific level of adherence to intervention).[6 316] Other ambiguous labels such as "modified intention to treat" are also variably defined from one trial to another.[314]

In addition to defining the analysis population, it is necessary to address the problem of missing data in the protocol. Most trials have some degree of missing data,[317 318] which can introduce bias depending on the pattern of "missingness" (eg, not missing at random). Strategies to maximise follow-up and prevent missing data, as well as the recording of reasons for missing data, are thus important to develop and document (Item 18b).[152]

The protocol should also state how missing data will be handled in the analysis and detail any planned methods to impute (estimate) missing outcome data, including which variables will be used in the imputation process (if applicable).[152] Different statistical approaches can lead to different results and conclusions,[317 319] but one study found that only 23% of trial protocols specified the planned statistical methods to account for missing data.[6]

Imputation of missing data allows the analysis to conform to intention to treat analysis but requires strong assumptions that are untestable and may be hard to justify.[152 318 320 321] Methods of multiple imputation are more complex but are widely preferred to single imputation methods (eg, last observation carried forward; baseline observation carried forward), as the latter introduce greater bias and produce confidence intervals that are too narrow.[152 320-322] Specific issues arise when outcome data are missing for crossover or cluster randomised trials.[323] Finally, sensitivity analyses are highly recommended to assess the robustness of trial results under different methods of handling missing data.[152 324]

Section 3d: Methods—monitoring

Data monitoring—formal committee

Item 21a: Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed

Explanation

For some trials, there are important reasons for periodic inspection of the accumulating outcome data by study group. In principle, a trial should be modified or discontinued when the accumulated data have sufficiently disturbed the clinical equipoise that justified the initiation of the trial. Data monitoring can also inform aspects of trial conduct, such as recruitment, and identify the need to make adjustments.

Example

"Appendix 3. Charter and responsibilities of the Data Monitoring Committee

A Data Monitoring Committee (DMC) has been established. The DMC is independent of the study organisers. During the period of recruitment to the study, interim analyses will be supplied, in strict confidence, to the DMC, together with any other analyses that the committee may request. This may include analyses of data from other comparable trials. In the light of these interim analyses, the DMC will advise the TSC [trial steering committee] if, in its view:

a) the active intervention has been proved, beyond reasonable doubt*, to be different from the control (standard management) for all or some types of participants, and

b) the evidence on the economic outcomes is sufficient to guide a decision from health care providers regarding recommendation of early lens extraction for PACG [primary angle closure glaucoma].

The TSC can then decide whether or not to modify intake to the trial. Unless this happens, however, the TSC, PMG [project management group], clinical collaborators and study office staff (except those who supply the confidential analyses) will remain ignorant of the interim results.

The frequency of interim analyses will depend on the judgement of the Chair of the DMC, in consultation with the TSC. However, we anticipate that there might be three interim analyses and one final analysis.

The Chair is Mr *D.G.-H.*, with Dr *D.C.*, and Professor *B.D.* Terms of reference for the DMC are available on request from the EAGLE [Effectiveness in Angle Closure Glaucoma of Lens Extraction] study office.

*Appropriate criteria for proof beyond reasonable doubt cannot be specified precisely. A difference of at least three standard deviation [sic] in the interim analysis of a major endpoint may be needed to justify halting, or modifying, such a study prematurely.[reference]"[325]

The decision to have a data monitoring committee (DMC) will be influenced by local standards. While certain trials warrant some form of data monitoring, many do not need a formal committee,[326] such as trials with a short duration or known minimal risks. A DMC was described in 65% (98/150) of cancer trial protocols with time-to-event outcomes in Italy in 2000-5,[327] and in 17% (12/70) of protocols for Danish randomised trials approved in 1994-5.[6] About 40% of clinical trials registered on ClinicalTrials.gov from 2007-2010 reported having a DMC.[328] The protocol should either state that there will be a DMC and provide further details, as discussed below, or indicate that there will not be a DMC, preferably with reasons.

When formal data monitoring is performed, it is often done by a DMC consisting of members from a variety of disciplines.[254 329] The primary role of a DMC is to periodically review the accumulating data and determine if a trial should be modified or discontinued. The DMC does not usually have executive power; rather, it communicates the outcome of its deliberations to the trial steering committee or sponsor.

Independence, in particular from the sponsor and trial investigators, is a key characteristic of the DMC and can be broadly defined as the committee comprising members who are "completely uninvolved in the running of the trial and who cannot be unfairly influenced (either directly

or indirectly) by people, or institutions, involved in the trial."[254] DMC members are usually required to declare any competing interests (Item 28). Among the 12 trial protocols that described a DMC and were approved in Denmark in 1994-5,[6] four explicitly stated that the DMC was independent from the sponsor and investigators; three had non-independent DMCs; and independence was unclear for the remaining five protocols.

The protocol should name the chair and members of the DMC. If the members are not yet known, the protocol can indicate the intended size and characteristics of the membership until further details are available. The protocol should also indicate the DMC's roles and responsibilities, planned method of functioning, and degree of independence from those conducting, sponsoring, or funding the trial.[254 330 331] A charter is recommended for detailing this information[331]; if this charter is not appended to the protocol, the protocol should indicate whether a charter exists or will be developed, and if so, where it can be accessed.

Data monitoring—interim analysis

Item 21b: Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial

Example

"Premature termination of the study

An interim-analysis is performed on the primary endpoint when 50% of patients have been randomised and have completed the 6 months follow-up. The interim-analysis is performed by an independent statistician, blinded for the treatment allocation. The statistician will report to the independent DSMC [data and safety monitoring committee]. The DSMC will have unblinded access to all data and will discuss the results of the interim-analysis with the steering committee in a joint meeting. The steering committee decides on the continuation of the trial and will report to the central ethics committee. The Peto approach is used: the trial will be ended using symmetric stopping boundaries at $P < 0.001$ [reference]. The trial will not be stopped in case of futility, unless the DSMC during the course of safety monitoring advices [sic] otherwise. In this case DSMC will discuss potential stopping for futility with the trial steering committee."[332]

Explanation

Interim analyses can be conducted as part of an adaptive trial design to formally monitor the accumulating data in clinical trials. They are generally performed in trials that have a DMC, longer duration of recruitment, and potentially serious outcomes. Interim analyses were described in 71% (106/150) of cancer trial protocols with time-to-event outcomes in Italy in 2000-5,[327] and in 19% (13/70) of protocols for Danish randomised trials approved in 1994-5.[6] The results of these analyses, along with non-statistical criteria, can be part of a stopping guideline that helps inform whether the trial should be continued, modified, or halted earlier than intended for benefit, harm, or futility. Criteria for stopping for harm are often different from those for benefit and might not employ a formal statistical criterion.[333] Stopping for futility occurs in instances where, if the study were to continue, it is unlikely that an important effect would be seen (ie, low chance of rejecting null hypothesis). Multiple analyses of the accumulating data increase the risk of a false positive (type I) error, and various statistical strategies have been developed to compensate for this inflated risk.[254 333-335] Aside from informing stopping guidelines, prespecified interim analyses can be used for other trial adaptations such as sample size re-estimation, alteration to the proportion of participants allocated to each study group, and changes to eligibility criteria.[111]

A complete description of any interim analysis plan, even if it is only to be performed at the request of an oversight body (eg, DMC), should be provided in the protocol—including the statistical methods, who will perform the analyses, and when they will be conducted (timing and indications). If applicable, details should also be provided about the decision criteria—statistical or other—that will be adopted to judge the interim results as part of a guideline for early stopping or other adaptations. Among 86 protocols for randomised trials with a time-to-event cancer outcome that proposed efficacy interim analyses, all stated the planned timing of the analyses, 91% specified the overall reason to be used for stopping (eg, superiority, futility), and 94% detailed the statistical approach.[327]

In addition, it is important to state who will see the outcome data while the trial is ongoing, whether these individuals will remain blinded (masked) to study groups, and how the integrity of the trial implementation will be protected (eg, maintaining blinding) when any adaptations to the trial are made. A third of protocols for industry initiated randomised trials receiving Danish ethics approval in 1994-95 stated that the sponsor had access to accumulating trial data, which can introduce potential bias due to competing interests.[10] Finally, the protocol should specify who has the ultimate authority to stop or modify the trial—eg, the principal investigator, trial steering committee, or sponsor.

Harms

Item 22: Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct

Explanation

Evaluation of harms has a key role in monitoring the condition of participants during a trial and in enabling appropriate management of adverse events. Documentation of trial related adverse events also informs clinical practice and the conduct of ongoing and future studies. We use the term "harms" instead of "safety" to better reflect the negative effects of interventions.[300] An adverse event refers to an untoward occurrence during the trial, which may or may not be causally related to the intervention or other aspects of trial participation.[300 336] This definition includes unfavourable changes in symptoms, signs, laboratory values, or health conditions. In the context of clinical trials, it can be difficult to attribute causation for a given adverse event. An adverse effect is a type of adverse event that can be attributed to the intervention.

Harms can be specified as primary or secondary outcomes (Item 12) or can be assessed as part of routine monitoring. To the extent possible, distinctions should be made between adverse events that are anticipated versus unanticipated, and solicited versus unsolicited, because expectation can influence the number and perceived severity of recorded events.

Example

"Secondary outcomes

. . . In our study an adverse event will be defined as any untoward medical occurrence in a subject without regard to the possibility of a causal relationship. Adverse events will be collected after the subject has provided consent and enrolled in the study. If a subject experiences an adverse event after the informed consent document is signed (entry) but the subject has not started to receive study intervention, the event will be reported as not related to study drug. All adverse events occurring after entry into the study and until hospital discharge will be recorded. An adverse event that meets the criteria for a serious adverse event (SAE) between study enrollment and hospital discharge will be reported to the local IRB [institutional review board] as an SAE. If haloperidol is discontinued as a result of an adverse event, study personnel will document the circumstances and data leading to discontinuation of treatment. A serious adverse event for this study is any untoward medical occurrence that is believed by the investigators to be causally related to study-drug and results in any of the following: Life-threatening condition (that is, immediate risk of death); severe or permanent disability, prolonged hospitalization, or a significant hazard as determined by the data safety monitoring board. Serious adverse events occurring after a subject is discontinued from the study will NOT be reported unless the investigators feels that the event may have been caused by the study drug or a protocol procedure. Investigators will determine relatedness of an event to study drug based on a temporal relationship to the study drug, as well as whether the event is unexpected or unexplained given the subject's clinical course, previous medical conditions, and concomitant medications.

. . . The study will monitor for the following movement-related adverse effects daily through patient examination and chart review: dystonia, akathisia, pseudoparkinsonism, akinesia, and neuroleptic malignant syndrome. Study personnel will use the Simpson-Angus [reference] and Barnes Akathisia [reference] scales to monitor movement-related effects.

. . .

For secondary outcomes, binary measures, eg mortality and complications, logistic regression will be used to test the intervention effect, controlling for covariates when appropriate . . ."[266]

For example, providing statements in the informed consent process about the possibility of a particular adverse effect or using structured, as opposed to open ended, questionnaires for data collection, can increase the reporting of specific events ("priming").[269 337-339] The timeframe for recording adverse events can also affect the type of data obtained.[340 341]

The protocol should describe the procedures for and frequency of harms data collection, the overall surveillance timeframe, any instruments to be used, and their validity and reliability, if known. Substantial discrepancies have been observed between protocol specified plans for adverse event collection and reporting, and what is described in final publications.[5] Although trials are often not powered to detect important differences in rates of uncommon adverse events, it is also important to describe plans for data analysis, including formal hypothesis testing or descriptive statistics.[300 342]

Finally, the protocol should address the reporting of harms to relevant groups (eg, sponsor, research ethics committee/institutional review board, data monitoring committee, regulatory agency), which is an important process that is subject to local regulation.[343] Key considerations include the severity of the adverse event, determination of potential causality, and whether it represents an unexpected or anticipated event. For multicentre studies, procedures and timing should be outlined for central collection, evaluation, and reporting of pooled harms data.

Example

"11.4 Data Monitoring and Quality Assurance

Through the combination of our web-based, instantaneous electronic validation, the DCC's [data coordinating centre] daily visual cross-validation of the data for complex errors, and regular on-site monitoring, the quality and completeness of the data will be reflective of the state of the art in clinical trials.

Both the European and US DCCs will conduct monitoring of source documents via fax at all enrolling ARUBA [A Randomised trial of Unruptured Brain Arteriovenous malformations] sites and will conduct at least one on-site monitoring visit per year over the course of the study at 100% of clinical sites (with repeat visits to sites where performance is a concern). Monitoring of European study sites will be assured by the European Coordinating Center (Paris). The primary objectives of the DCC during the on-site visits are to educate, support and solve problems. The monitors will discuss the protocol in detail and identify and clarify any areas of weakness. At the start of the trial, the monitors will conduct a tutorial on the web-based data entry system. The coordinators will practice entering data so that the monitors can confirm that the coordinators are proficient in all aspects of data entry, query response, and communication with the DCC. They will audit the overall quality and completeness of the data, examine source documents, interview investigators and coordinators, and confirm that the clinical center has complied with the requirements of the protocol. The monitors will verify that all adverse events were documented in the correct format, and are consistent with protocol definition.

The monitors will review the source documents as needed, to determine whether the data reported in the Web-based system are complete and accurate. Source documents are defined as medical charts, associated reports and records including initial hospital admission report . . .

The monitors will confirm that the regulatory binder is complete and that all associated documents are up to date. The regulatory binder should include the protocol and informed consent (all revisions), IRB [institutional review board] approvals for all of the above documents, IRB correspondence, case report forms, investigator's agreements . . .

Scheduling monitoring visits will be a function of patient enrollment, site status and other commitments. The DCC will notify the site in writing at least three weeks prior to a scheduled visit. The investigators must be available to meet with the monitors. Although notification of the visits will include the list of patients scheduled to be reviewed, the monitors reserve the right to review additional ARUBA patients.

If a problem is identified during the visit (ie, poor communication with the DCC, inadequate or insufficient staff to conduct the study, missing study documents) the monitor will assist the site in resolving the issues. Some issues may require input from the Operations Committee, Steering Committee or one of the principal investigators.

The focus of the visit/electronic monitoring will be on source document review and confirmation of adverse events. The monitor will verify the following variables for all patients: initials, date of birth, sex, signed informed consent, eligibility criteria, date of randomization, treatment assignment, adverse events, and endpoints . . ."[313]

Auditing

Item 23: Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor

Explanation

Auditing involves periodic independent review of core trial processes and documents. It is distinct from routine day-to-day measures to promote data quality (Items 18a and 19). Auditing is intended to preserve the integrity of the trial by independently verifying a variety of processes and prompting corrective action if necessary. The processes reviewed can relate to participant enrolment, consent, eligibility, and allocation to study groups; adherence to trial interventions and policies to protect

participants, including reporting of harms (Item 22); and completeness, accuracy, and timeliness of data collection. In addition, an audit can verify adherence to applicable policies such as the International Conference on Harmonisation *Good Clinical Practice* and regulatory agency guidelines.[160]

In multicentre trials, auditing is usually considered both overall and for each recruiting centre. Audits can be done by exploring the trial dataset or performing site visits. Audits might be initially conducted across all sites, and subsequently conducted using a risk based approach that focuses, for example, on sites that have the highest enrolment rates, large numbers of withdrawals, or atypical (low or high) numbers of reported adverse events.

If auditing is planned, the procedures and anticipated frequency should be outlined in the protocol, including a description of the personnel involved and their degree of independence from the trial investigators and sponsor. If procedures are further detailed elsewhere (eg, audit manual), then the protocol should reference where the full details can be obtained.

Section 4: Ethics and dissemination

Research ethics approval

Item 24: Plans for seeking research ethics committee/institutional review board (REC/IRB) approval

Example

"This protocol and the template informed consent forms contained in Appendix II will be reviewed and approved by the sponsor and the applicable IRBs/ECs [institutional review boards/ethical committees] with respect to scientific content and compliance with applicable research and human subjects regulations. . . .

The protocol, site-specific informed consent forms (local language and English versions), participant education and recruitment materials, and other requested documents—and any subsequent modifications — also will be reviewed and approved by the ethical review bodies. . .

Subsequent to initial review and approval, the responsible local Institutional Review Boards/Ethical Committees (IRBs/ECs) will review the protocol at least annually. The Investigator will make safety and progress reports to the IRBs/ECs at least annually and within three months of study termination or completion at his/her site. These reports will include the total number of participants enrolled . . . and summaries of each DSMB [data safety and monitoring board] review of safety and/or efficacy."[287]

Explanation

A universal requirement for the ethical conduct of clinical research is the review and approval of the research protocol by qualified individuals who are not associated with the research team and have no disqualifying competing interests as reviewers.[1] The review is typically conducted by a formal REC/IRB in accordance with jurisdictional policy. Despite the importance of ethics review, approval by a REC/IRB is not always obtained. Among 767 trials published in leading general medical journals from 1993-95, 37 authors (5%) disclosed that such approval had not been sought for their trials.[344] The protocol should document where approval has been obtained, or outline plans to seek such approval.

Protocol amendments

Item 25: Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)

Example

"13.10 Modification of the Protocol

Any modifications to the protocol which may impact on the conduct of the study, potential benefit of the patient or may affect patient safety, including changes of study objectives, study design, patient population, sample sizes, study procedures, or significant administrative aspects will require a formal amendment to the protocol. Such amendment will be agreed upon by BCIRG [Breast Cancer International Research Group] and Aventis, and approved by the Ethics Committee/IRB [institutional review board] prior to implementation and notified to the health authorities in accordance with local regulations.

Administrative changes of the protocol are minor corrections and/or clarifications that have no effect on the way the study is to be conducted. These administrative changes will be agreed upon by BCIRG and Aventis, and will be documented in a memorandum. The Ethics Committee/IRB may be notified of administrative changes at the discretion of BCIRG."[345]

Explanation

After initial ethics approval, about half of trials have subsequent protocol amendments submitted to the REC/IRB.[125 346 347] While some amendments may be unavoidable, a study of pharmaceutical industry trials found that according to the sponsors, a third of amendments could have been prevented with greater attention to key issues during protocol development.[346] Substantive amendments can generate challenges to data analysis and interpretation if they occur part way through the trial (eg, changes in eligibility criteria),[348] and can introduce bias if the changes are made based on the trial data.[173-176] The implementation and communication of amendments are also burdensome and potentially costly.[346]

Numerous studies have revealed substantive changes between prespecified methods (eg, as stated in approved protocols, registries, or regulatory agency submissions) and those described in trial publications, including changes to primary outcomes,[12 172-176] sample size calculations,[6] eligibility criteria,[125 133 134] as well as methods of allocation concealment,[2] blinding,[3] and statistical analysis.[6-8 174] These substantive modifications are rarely acknowledged in the final trial reports, providing an inaccurate impression of trial integrity.

It is important that substantive protocol amendments be reviewed by an independent party, such as the REC/IRB, and transparently described in trial reports. The notion of "substantive" is variably defined by authorities, but in general refers to a protocol amendment that can affect the safety of trial participants or the scientific validity, scope, or ethical rigour of the trial.[349 350] To reflect the degree of oversight for the trial and adherence to applicable regulation, the protocol should describe the process for making amendments, including who will be responsible for the decision to amend the protocol and how substantive changes will be communicated to relevant stakeholders (eg, REC/IRBs, trial registries, regulatory agencies). Version

control using protocol identifiers and dates (Item 3), as well as a list of amendments, can help to track the history of amendments and identify the most recent protocol version.

Consent or assent

Item 26a: Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)

Example

"...Trained Research Nurses will introduce the trial to patients who will be shown a video regarding the main aspects of the trial. Patients will also receive information sheets. Research Nurses will discuss the trial with patients in light of the information provided in the video and information sheets. Patients will then be able to have an informed discussion with the participating consultant. Research Nurses will obtain written consent from patients willing to participate in the trial. Information sheets and consent forms are provided for all parents involved in the trial however these have been amended accordingly in order to provide separate information sheets and consent form [sic] which are suitable for children and teenagers. All information sheets, consent forms and the video transcript have been translated into Bengali, Punjabi, Gujarati, and Urdu. There are also separate information sheets and consent forms for the cohort group."[351]

Explanation

The notion of acquiring informed consent involves the presentation of comprehensible information about the research to potential participants, confirmation that they understand the research, and assurance that their agreement to participate is voluntary. The process typically involves discussion between the potential participant and an individual knowledgeable about the research; the presentation of written material (eg, information leaflet or consent document); and the opportunity for potential participants to ask questions. Surveys of trial investigators reveal that appropriate informed consent is not always obtained.[344 352]

The content, quantity, and mode of delivery of consent information can affect trial recruitment, participant comprehension, anxiety, retention rates, and recruitment costs.[68 114 218 292 353-355] We recommend that a model consent or assent form be provided as a protocol appendix (Item 32). Assent represents a minor's affirmative agreement to participate in the trial, which typically involves signing a document that provides age appropriate information about the study.

The protocol should include details of the consent process as well as the status, experience, and training (if applicable) of the research team members who will conduct it. In paediatric research, regulations may stipulate obtaining affirmative assent for participation from children above a certain age.[356] The protocol should then describe how pertinent information will be provided to potential participants and how their understanding and assent will be ascertained. When potential participants lack decisional capacity for reasons other than young age (eg, mental status), and proxy consent can be obtained from a legally-authorised representative, the protocol should describe who will determine an individual's decisional capacity, whether a formal capacity instrument will be utilised, and how the individual's informed agreement to continue participation will be secured should they regain decisional capacity. For certain trials, such as cluster randomised trials, it may not be possible to acquire individual informed consent from participants before randomisation, and the consent process may be modified or waived. An explanation should be provided in the protocol in these instances.[357]

Consent or assent—ancillary studies

Item 26b: Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable

Example

"6.4.1. Samples for Biorepositories

Additional biological samples will be obtained to be stored for use in future studies of the pathobiology of FSGS [focal segmental glomerulosclerosis]. A materials consent will be obtained to specifically address the collection of these ... urine, serum and plasma specimens ...

14.3.4. Instructions for Preparation of Requests for an Ancillary Study

... A signed consent must be obtained from every participant in the ancillary study, if the data collection/request is not covered in the original informed consent process for the main FSGS Clinical Trial.

...

A copy of the IRB [institutional review board] letter for the ancillary study should be sent to the DCC [data coordinating centre]. If a separate consent form is required for the ancillary study, a copy of the signed ancillary study consent form for each study participant must be included in the FSGS-CT [clinical trial] record. A data file tracking all signed ancillary consent forms must be maintained by the ancillary study and an electronic copy of that file must be delivered to the FSGS-CT DCC."[267]

Explanation

Ancillary studies involve the collection or derivation of data for purposes that are separate from the main trial. The acquisition and storage of data and biological specimens for ancillary studies is increasingly common in the context of clinical trials (Item 33). Specimens may be used for a specified subset of studies or for submission to biorepositories for future specified or unspecified research.

Ancillary studies have additional processes and considerations relating to consent, which should be detailed in the protocol. Guidance for the creation of a simplified informed consent document for biobanking is available.[358] Participants can be given several options to consider with respect to their participation in ancillary research: consent for the use of their data and specimens in specified protocols; consent for use in future research unrelated to the clinical condition under study; consent for submission to an unrelated biorepository; and consent to be contacted by trial investigators for further informational and consent-related purposes. This is commonly referred to as tiered consent. Participants should also be informed about whether their withdrawal from the ancillary research is possible (eg, the data and specimens are coded and identifiable); what withdrawal means in this context (eg, used specimens and data derived from them cannot be withdrawn); and what information derived from the specimen related research will be provided to them, if any.

Confidentiality

Item 27: How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial

Example

"8.5 Confidentiality

All study-related information will be stored securely at the study site. All participant information will be stored in locked file cabinets in areas with limited access. All laboratory specimens, reports, data collection, process, and administrative forms will be identified by a coded ID [identification] number only to maintain participant confidentiality. All records that contain names or other personal identifiers, such as locator forms and informed consent forms, will be stored separately from study records identified by code number. All local databases will be secured with password-protected access systems. Forms, lists, logbooks, appointment books, and any other listings that link participant ID numbers to other identifying information will be stored in a separate, locked file in an area with limited access.

All HIV test results will be kept strictly confidential, all counseling and blood draws will be conducted in private rooms, and study staff will be required to sign agreements to preserve the confidentiality of all participants. Study staff will never inform network members of the serostatus of other members of their group, but counselors will provide general messages about the prevalence of HIV in the study population in the interests of emphasizing harm reduction.

Participants' study information will not be released outside of the study without the written permission of the participant, except as necessary for monitoring by NIAID [National Institute of Allergy and Infectious Diseases] and/or its contractors . . . representatives of the HPTN CORE [HIV Prevention Trials Network Coordinating and Operations Center] . . . and US or in-country government and regulatory authorities."[359]

Explanation

Personal information about participants is acquired during the process of trial recruitment, eligibility screening, and data collection. Much of this information consists of private details over which people customarily wish to maintain control, such as their health status, personal genotype, and social and family history.

The protocol should describe the means whereby personal information is collected, kept secure, and maintained. In general, this involves: 1) the creation of coded, depersonalised data where the participant's identifying information is replaced by an unrelated sequence of characters; 2) secure maintenance of the data and the linking code in separate locations using encrypted digital files within password protected folders and storage media; and 3) limiting access to the minimum number of individuals necessary for quality control, audit, and analysis. The protocol should also describe how the confidentiality of data will be preserved when the data are transmitted to sponsors and coinvestigators (eg, virtual private network internet transmission).

Declaration of interests

Item 28: Financial and other competing interests for principal investigators for the overall trial and each study site

Example

"*PS*:

1. Was the Principal Investigator of the second International Stroke Trial (IST-2) to evaluate a neuroprotective compound (619c89). . .

2. Has received lecture fees and travel expenses from Bayer and from Boehringer Ingelheim for lectures given at international conferences.

3. He serves on the Independent Data Monitoring and Safety Board of the RELY trial, funded by Boehringer Ingelheim and receives attendance fees and travel expenses for attending board meetings.

4. He does not have any paid consultancies with pharmaceutical companies, and is not a member of the Speaker's Panel of any company.

KBS:

Received an honorarium for a lecture from Boehringer Ingelheim and had costs for participating in scientific meetings reimbursed. . ." [124]

Explanation

Competing interests, or conflicts of interest, exist when there is potential for divergence between an individual's or institution's private interests and their responsibilities to scientific and publishing activities.[360] More positive outcomes, larger treatment effect sizes, and more favourable interpretation of results have been found in clinical trials with pharmaceutical industry sponsorship (Item 4)[27 36-38 42] and investigators who have declared competing interests,[57 60] compared to those without such interests. Although competing interests are most often associated with drug and device industries, they may exist with support from or affiliation with government agencies, charities, not for profit organisations, and professional and civic organisations.

Competing interests do not in themselves imply wrongdoing. Their disclosure and regular updating enables appropriate management plans to be developed and implemented, and facilitates transparent assessment of the potential for bias.

Many trials and non-industry sponsors have a conflict of interest policy for their investigators, and checklists are available to guide potential interests that should be disclosed and regularly updated by trial investigators.[361 362] Types of financial ties include salary support or grants; ownership of stock or options; honorariums (eg, for advice, authorship, or public speaking); paid consultancy or service on advisory boards and medical education companies; and receipt of patents or patents pending. Non-financial competing interests include academic commitments; personal or professional relationships; and political, religious, or other affiliations with special interests or advocacy positions.

Access to data

Item 29: Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators

Explanation

The validity of results from interventional trials can be verified only by individuals who have full access to the complete final dataset. For some multicentre trials, only

Example

"12.10.1 Intra-Study Data Sharing
The Data Management Coordinating Center will oversee the intra-study data sharing process, with input from the Data Management Subcommittee.
All Principal Investigators (both US and host country) will be given access to the cleaned data sets. Project data sets will be housed on the Project Accept Web site and/or the file transfer protocol site created for the study, and all data sets will be password protected. Project Principal Investigators will have direct access to their own site's data sets, and will have access to other sites data by request. To ensure confidentiality, data dispersed to project team members will be blinded of any identifying participant information."[113]

the steering group has access to the full trial dataset in order to ensure that the overall results are not disclosed by an individual study site prior to the main publication. Many of these trials will allow site investigators to access the full dataset if a formal request describing their plans is approved by the steering group. The World Medical Association supports the principle that trial investigators retain the right to access data.[363] However, among protocols of industry initiated randomised trials published in 2008-9 in the *Lancet* or approved in 2004 by a Danish ethics committee, 30-39% stated that the sponsor owned the data while 0-3% stated that principal investigators had access to all trial data.[10 364] Similar constraints were found in Danish trial protocols from 1994-5.[10]

The protocol should identify the individuals involved in the trial who will have access to the full dataset. Any restrictions in access for trial investigators should also be explicitly described.

Ancillary and post-trial care

Item 30: Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation

Explanation

The provision of ancillary care refers to the provision of care beyond that immediately required for the proper and safe conduct of the trial, and the treatment of immediate adverse events related to trial procedures. It is generally agreed that trial sponsors and investigators should plan to provide care for participants' healthcare needs that arise as a direct consequence of trial participation (eg, intervention related harms). It is also important to consider whether care should be provided for certain ancillary needs that may otherwise arise during trial participation. Provision of care for ancillary needs reflects the fact that participants implicitly, but unavoidably, entrust certain aspects of their health to the research team. The scope of entrustment will vary depending on the nature of the trial (eg, setting, health condition under study, investigations performed).[366] Additional factors that influence the strength of the claim to ancillary care include participants' vulnerabilities; uncompensated burdens and harms; the intensity and duration of the participant-researcher relationship; and the degree to which participants are uniquely dependent on the research team for health care.[367]

The Declaration of Helsinki states that "the protocol should describe arrangements for post-study access by

Examples

"Patients that are enrolled into the study are covered by indemnity for negligent harm through the standard NHS [National Health Service] Indemnity arrangements. The University of Sheffield has insurance to cover for non-negligent harm associated with the protocol . . . This will include cover for additional health care, compensation or damages whether awarded voluntarily by the Sponsor, or by claims pursued through the courts. Incidences judged to arise from negligence (including those due to major protocol violations) will not be covered by study insurance policies. The liability of the manufacturer of IL1RA (Amgen Corporation) is strictly limited to those claims arising from faulty manufacturing of the commercial product and not to any aspects of the conduct of the study."[145]

"13.6 Access to Effective Products
Should this study provide evidence of the effectiveness of TDF [tenofovir disoproxil fumarate], FTC [emtricitabine]/TDF and/or tenofovir 1% gel in preventing HIV infection, it will be critical to provide access to the effective product(s) to study participants, their communities, and the worldwide population at risk for HIV infection in a timely manner. In preparation for this study, discussions have begun with Gilead Sciences, Inc. and CONRAD [Contraceptive Research and Development Organization] to ensure such access. Considerations under discussion include licensing agreements and preferred pricing arrangements for the study communities and other resource-poor settings.
While this study is ongoing, the MTN [Microbicide Trials Network] will continue these discussions. In addition, discussions will be initiated with other public and private funding sources such as the WHO, UNAIDS, Gates Foundation, and appropriate site government agencies that may be able to purchase product supplies in bulk and offer them at low or no cost to the study communities and other resource-poor communities most in need of the product(s). Operations and marketing research also may be conducted to determine how best to package and distribute the products, and maximize their acceptability and use, in at-risk populations."[365]

study participants to interventions identified as beneficial in the study or access to other appropriate care or benefits."[1] This principle is particularly applicable—and controversial—when research enabling the development and regulatory approval of interventions is performed in countries where subsequent access to the interventions is limited by cost or lack of availability.[368]

The protocol should describe any plans to provide or pay for ancillary care during the trial and identify any interventions, benefits, or other care that the sponsor will continue to provide to participants and host communities after the trial is completed.[369] Any plans to compensate participants for trial related harms should also be outlined.

Dissemination policy—trial results

Item 31a: Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions

Explanation

A fundamental ethical principle in clinical trials is that the potential risks incurred by study participants should be balanced by the benefit of contributing to publicly available knowledge.[371] Unfortunately, about half of

Example

"XII. Publication Policy
The Publications subcommittee will review all publications following the guidelines given below and report its recommendations to the Steering Committee.

A. Data analysis and release of results
The scientific integrity of the project requires that the data from all BEST [Beta-Blocker Evaluation of Survival Trial] sites be analyzed study-wide and reported as such. Thus, an individual center is not expected to report the data collected from its center alone . . . all presentations and publications are expected to protect the integrity of the major objective(s) of the study; data that break the blind will not be presented prior to the release of mainline results. Recommendations as to the timing of presentation of such endpoint data and the meetings at which they might be presented will be given by the Steering Committee.

B. Review process
Each paper or abstract, as described below, must be submitted to the appropriate Subcommittee for review of its appropriateness and scientific merit prior to submission. The Subcommittee may recommend changes to the authors and will finally submit its recommendations to the Steering Committee for approval.

C. Primary outcome papers
The primary outcome papers of BEST are papers that present outcome data . . . The determination of whether or not a particular analysis represents a primary outcome will be made by the Steering Committee on the recommendation of the Publications Subcommittee . . .

D. Other study papers, abstracts and presentations
All studies other than those designated as "Primary Outcome" fall within this category . . . All papers and abstracts must be approved by the Publications Committee before they are submitted.

It is possible that in certain instances BEST may be asked to contribute papers to workshops, symposia, volumes, etc. The individuals to work on such requests should be appointed by the Executive Committee, but where time permits, a proposal will be circulated soliciting other participants as in the case of other study papers as described in the Application Review Process.

XIII. Close-out Procedures
BEST may terminate at the planned target of 1.5 years after the last participant has been randomized, or at an earlier or later date if the circumstances warrant . . . Regardless of the timing and circumstances of the end of the study, close-out will proceed in two stages:
- Interim period for analysis and documentation of study results.
- Debriefing of participants and dissemination of study results.

A. Interim
Every attempt will be made to reduce to an absolute minimum the interval between the completion of data collection and the release of the study results. We expect to take about 3 to 4 months to compile the final results paper for an appropriate journal.

B. Reporting of study results
The study results will be released to the participating physicians, referring physicians, patients and the general medical community."[370]

clinical trials remain unpublished.[80 83] Trials with statistically non-significant results or industry funding are more prone to non-publication,[36 38 80-83] although government funded trials are also susceptible.[81] When published, trials with non-significant results often have a longer delay to publication.[80 83] Overall, the medical literature represents a biased subset of existing data, potentially leading to overestimation of benefits, underestimation of harms, and a detrimental impact on patient care and research.[80 372-377]

Although peer reviewers can be biased in favour of positive findings,[378] lack of publication appears to be primarily due to trial investigators or sponsors failing to submit negative or null results, rather than journals rejecting them.[80 379] A plan to disseminate trial results to key stakeholders should be outlined in the protocol, including a process and timeframe for approving and submitting reports for dissemination (eg, via journal publication, trial registry, trial website), and an explicit statement that the results will be disseminated regardless of the magnitude or direction of effect.

Furthermore, any conditions relating to the investigators' right to publish or present trial results should be explicitly described. Publication restrictions have been imposed by various groups, including industry sponsors or the trial steering group (eg, to maintain the integrity of the overall dataset).[10 380] These restrictions are sometimes not described in the protocol but rather in separate publication agreements.[10] However, as they can interfere with the ethical responsibility of investigators and sponsors to disseminate trial results in an unbiased and timely manner,[38 381-384] any restrictions should be disclosed in the protocol for review by REC/IRBs, funders, and other stakeholders. A review of industry initiated randomised trial protocols approved in Denmark in 1994-95 revealed that 91% had publication restrictions imposed by sponsors; similar constraints were noted for protocols approved in 2004.[10]

Dissemination policy—authorship

Item 31b: Authorship eligibility guidelines and any intended use of professional writers

Example

"17.4. Assignment of Writing Committees
Topics suggested for presentation or publication will be circulated to the PIs [principal investigators] of the CCCs [core coordinating centers], the DCC [data coordinating centre], Core Lab and the NIH [National Institutes of Health]. These groups are requested to suggest and justify names for authors to be reviewed by the PC [publications committee]. . . If a topic is suggested by a participant of the FSGS-CT [focal segmental glomerulosclerosis—clinical trial], the writing committee will be formed as just described except that the person making the suggestion may be considered as the lead author. The PI of an ancillary study should be considered for lead author of material derived from this study. Disputes regarding authorship will be settled by the Study Chair after consultation with the Chair of the PC . . .

17.5. Reports of the FSGS-CT: Classes of Reports
There are three classes of reports of the FSGS-CT:

A. Reports of the major outcomes of the Study.

B. Reports addressing in detail one aspect of the FSGS-CT, but in which the data are derived from the entire study.

C. Reports of data derived from a subset of centers by members of the FSGS-CT, (eg, sub-studies or ancillary studies), or reports of investigations initiated outside of the FSGS-CT, but using data or samples collected by the FSGS-CT. . .

17.6. Authorship Policy
The authors of FSGS publications will be listed as detailed below.

Type A publications:

abstracts: from the FSGS Clinical Trial Group[x], presented by XXXX.

papers: from the FSGS Clinical Trial Group[x], prepared by XXXX.

[x]The FSGS participant box, detailed below, must be included in these papers. If a journal's publication policy does not allow authorship by a group, the authors will be listed first as in Type B publications.

Type B publications:

. . .

17.7. Authorship: Professional Participants Listing in the FSGS Participant Box
The FSGS participant box will list all professionals that have participated in the FSGS-CT for a minimum of one year."[267]

Explanation

Substantive contributions to the design, conduct, interpretation, and reporting of a clinical trial are recognised through the granting of authorship on the final trial report. Authorship guidelines in the protocol are intended to help enhance transparency and avoid disputes or misunderstanding after trial completion. These guidelines should define criteria for individually named authors or group authorship.[385]

Individuals who fulfil authorship criteria should not remain hidden (ghost authorship) and should have final authority over manuscript content.[9 386 387] Similarly, those who do not fulfil such criteria should not be granted authorship (guest authorship).[386 388] The International Committee of Medical Journal Editors has defined authorship criteria for manuscripts submitted for publication,[389] although these criteria have reportedly been open to abuse.[390] If some protocol authors are not named authors of subsequent publications, their role in protocol design should at least be acknowledged in the published report. Among 44 protocols of industry initiated trials, 75% had evidence of ghost authorship when compared with corresponding journal publications.[9]

Professional medical writers are sometimes hired to improve clarity and structure in a trial report, and guidelines for ethical collaborative writing have been developed.[391 392] Because the drafting of text can influence how the study results and conclusions are portrayed, plans for the employment of writers and their funding source should be acknowledged in both protocols and trial reports.

Dissemination policy—reproducible research

Item 31c: Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code

Example

"**Data sharing statement** No later than 3 years after the collection of the 1-year postrandomisation interviews, we will deliver a completely deidentified data set to an appropriate data archive for sharing purposes."[393]

Explanation

Given the central role of protocols in enhancing transparency, reproducibility, and interpretation of trial results, there is a strong ethical and scientific imperative to ensure that full protocols are made publicly available.[24 394 395] High quality protocols contain relevant details on study design and conduct that are generally not available in journal publications or trial registries.[84 396] It is also important to make available the full study report, such as the "clinical study report" submitted to regulatory agencies by industry sponsors.[377 396-400] This detailed report provides the most comprehensive description of trial methods (including the full protocol) and all published and unpublished analyses. In addition, there have increasingly been calls to improve the availability of participant-level datasets and statistical code after journal publication to enable verification and replication of analyses, facilitate pooling with other studies, and accelerate research through open knowledge sharing.[372 401-406]

Avenues for providing access to full protocols include journals,[407 408] trial websites, and trial registries.[163] Several journals and funders support the sharing of participant level data,[405 409-411] while others routinely publish a statement regarding sharing of protocols, statistical codes, and datasets for all of their published research articles.[412 413] The protocol should indicate whether the trial protocol, full study report, anonymised participant level dataset, and statistical code for generating the results will be made publicly available; and if so, describe the timeframe and any other conditions for access.

Section 5: Appendices

Informed consent materials

Item 32: Model consent form and other related documentation given to participants and authorised surrogates

Example

"APPENDIX 7 SAMPLE PATIENT INFORMED CONSENT

Note: . . . Each Ethics Committee or Institutional Review Board will revise and adapt according to their own institution's guidelines.

MULTICENTER PHASE III RANDOMIZED TRIAL COMPARING DOXORUBICIN AND CYCLOPHOSPHAMIDE . . .

Study number: BCIRG 006 (TAX GMA 302)

Investigator name:

Address:

Consent Form:

This consent form is part of the informed consent process. It is designed to give you an idea of what this research study is about and what will happen to you if you choose to be in the study. . ." [345]

Explanation

The Declaration of Helsinki states that each potential trial participant must normally, at a minimum, be adequately informed about the purpose of the trial; potential benefits and risks; their right to refuse participation or to withdraw consent at any time; institutional affiliation and potential competing interests of the researcher; and sources of trial funding.[1] There are rare exceptions where deferred consent can be acceptable, such as trials involving unconscious patients in emergency situations.

Special attention is required to ensure that relevant information is provided and appropriate modes of delivery are used during the consent process (Item 26).[414] Consent and participant information forms are often written at a much higher reading level than is acceptable for the general population.[415] Depending on the nature of the trial, several different consent documents may be needed. For example, a paediatric trial may involve both parental permission and participant assent documents. For multicentre trials, a model or sample document is typically drafted for distribution to local investigators, who may then revise the document to comply with local requirements.

Biological specimens

Item 33: Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable

Example

"White Blood Cell and Plasma Collection Procedures

1.0 Objectives

1.1 To provide a resource for studies of early markers, etiology, and genetic risk factors for prostate cancer and other diseases.

2.0 Background

The Prostate Cancer Prevention Trial (PCPT) is a randomized double blind chemoprevention trial . . . Initial blood collection was specifically for the analysis of PSA [prostate specific antigen] and storage of serum . . . an additional blood collection will be carried out using anticoagulant so that plasma and white blood cells can be isolated. Plasma will allow the analysis of additional biomarkers . . . This DNA will be used (among other possible uses) for studies to investigate polymorphisms in genes which may influence prostate cancer risk . . .

The PCPT WBC [white blood cell] sample will be available to PCPT investigators as well as outside researchers who have important, timely hypotheses to test. Because the sample bank is a limited resource, proposals to use it will be evaluated in terms of scientific relevance, significance, and validity as well as the potential impact of the proposed study. The amount and type of material needed will also be considered and the efficient use of material will be required. Strict confidentiality will be exercised and the information provided to investigators will not contain personal identifiers. When specific uses of the WBC samples are approved, the SWOG-9217 protocol will be amended. Participation in this research is not required for continued participation in the PCPT.

3.0 Methods

3.1 Because the original model consent form did not specifically address genetic studies, participants will be asked to sign an additional consent form to document their consent to the collection and submission of additional blood samples for storage and future testing (including genetic analysis).

3.2 Institutions will be asked to submit additional materials from participants who consent to the additional blood collection. The blood is to be collected, processed and shipped as described in the PCPT Study Manual.

3.3 NCI-Frederick Cancer Research Development Center (FCRDC) in Frederick, Maryland will serve as the processing, aliquotting and storage facility.

3.4 Upon arrival at FCRDC the blood will be pooled and centrifuged. Plasma will be separated into 5 x 1.8 ml aliquots and frozen . . .

3.5 All samples will be logged in and aliquots will be bar coded with a unique storage ID. These data will be electronically transmitted to the Statistical Center for verification.

3.6 The scientists who will carry out analyses on these materials will not have access to personal identifiers and will not be able to link the results of these tests to personal identifier information. No individual results will be presented in publications or other reports. . . .

3.7 Participants will not be informed on an individual basis of any results from these studies . . .

4.0 Sample analysis

4.1 Investigators planning to submit NIH [National Institutes of Health] grant applications must obtain approval for their study and specimen access from the PCPT Serum and Tissue Utilization Committee before submission of a grant proposal. Potential investigators will be required to submit a brief abstract and 1-4 page outline . . . This proposal will be circulated for review to members of the PCPT Serum and Tissue Utilization Committee and two ad hoc members having relevant expertise . . .

4.2 It is anticipated that proposals will be reviewed once a year . . . Approval by this group as well as appropriate Institutional Review Board approval from the investigator's institution will be required before release of samples."[416]

Explanation

Biological specimens (eg, biopsy tissue; blood for DNA extraction) obtained during the conduct of clinical trials can be stored in repositories—often designated as biobanks—for the current trial and future research. This process is usually governed by local regulation and has particular ethical considerations (Item 26b).

If the trial involves genetic or molecular analysis of biological specimens derived from humans, or if any specimens will be stored for future use (specified or unspecified), the protocol should describe details about specimen collection, storage, and evaluation, including the location of repositories. In addition, the protocol should state whether collected samples and associated participant related data will be de-identified or coded to protect participant confidentiality. If a repository is overseen by a named research ethics committee/institutional review board, then this information should also be provided.

Discussion

It is critical that every clinical trial has a complete and transparent protocol, which can then facilitate trial conduct and appraisal by communicating relevant information to key stakeholders. In response to observed deficiencies in protocol content, the SPIRIT Initiative has produced recommendations for minimum relevant protocol items to include in a protocol, published in the form of the SPIRIT 2013 Statement and this Explanation and Elaboration (E&E) paper.[14] The strengths that distinguish SPIRIT from other protocol guidance documents include its systematic and transparent development methods; participation of a wide range of key stakeholders; use of empirical evidence to support its recommendations; and availability of detailed guidance including model examples from protocols.

The overall aim of SPIRIT is to improve the completeness and transparency of trial protocols. The SPIRIT documents can serve as a practical resource for trial investigators and personnel to draft and understand the key elements of a protocol. In doing so, our vision is that the SPIRIT 2013 Statement and E&E paper will also facilitate and expedite the review of protocols by research ethics committees/institutional review boards, scientific review groups, and funders—for example, by reducing the number of avoidable queries to trial investigators regarding missing or unclear protocol information during the review process. Furthermore, improved protocol content would help facilitate the critical appraisal of final trial reports and results. Finally, several SPIRIT items correspond to items on the CONSORT 2010 checklist (Consolidated Standards of Reporting Trials),[417] which should facilitate the transition from the protocol to the final study report.

The next steps for the SPIRIT Initiative include an implementation strategy to encourage uptake of the SPIRIT 2013 Statement. The SPIRIT website (www.spirit-statement.org) will provide the latest resources and information on the initiative, including a list of supporters. We invite stakeholders to assist in the evaluation of the SPIRIT Statement and E&E paper by using the documents and providing feedback to inform future revisions. Through widespread uptake and support, the potential to improve the completeness and quality of trial protocols, as well as the efficiency of their review, can be fully realised.

We thank Raymond Daniel for his help with reference management and Jessica Kitchen for her work with manuscript formatting and identification of protocol examples. We also acknowledge GlaxoSmithKline for providing a sample of their trial protocols to serve as potential examples.

Competing interests: All authors have completed the ICJME unified declaration form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: JAB is employed by the Janssen Pharmaceutical Companies of Johnson & Johnson; KKJ was formerly employed by CIHR (Knowledge Translation Branch), and WRP is affiliated with the NCIC Clinical Trials Group. Trish Groves is deputy editor of *BMJ* and a member of the SPIRIT group but did not take part in the peer review and decision making process about this publication.

Contributors: AWC, JT, and DM conceived of the paper. All authors contributed to the drafting and revision of the manuscript, and approve the final version. AWC is the guarantor for the article.

Funding: The SPIRIT meetings were funded by the Canadian Institutes of Health Research (CIHR grant DET - 106068); National Cancer Institute of Canada (now Canadian Cancer Society Research Institute); and Canadian Agency for Drugs and Technologies in Health. CIHR has also funded ongoing dissemination activities (grant MET-117434). KKJ was formerly employed by CIHR (Knowledge Translation Branch), and WRP is affiliated with the NCIC Clinical Trials Group. The funders had no input into the

design and conduct of the project; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 World Medical Association. WMA Declaration of Helsinki—ethical principles for medical research involving human subjects. 2008. www.wma.net/en/30publications/10policies/b3/index.html.
2 Pildal J, Chan A-W, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005;330:1049.
3 Hróbjartsson A, Pildal J, Chan A-W, Haahr MT, Altman DG, Gøtzsche PC. Reporting on blinding in trial protocols and corresponding publications was often inadequate but rarely contradictory. *J Clin Epidemiol* 2009;62:967-73.
4 Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
5 Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J Clin Oncol* 2006;24:3933-8.
6 Chan A-W, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
7 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by the Lancet. *Lancet* 2008;372:201.
8 Hernández AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery* 2005;57:1244-53.
9 Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan A-W. Ghost authorship in industry-initiated randomised trials. *PLoS Med* 2007;4:e19.
10 Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan A-W. Constraints on publication rights in industry-initiated clinical trials. *JAMA* 2006;295:1645-6.
11 Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol* 2012;65:602-9.
12 Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. *BMJ* 2011;342:c7153.
13 Tetzlaff JM, Chan A-W, Kitchen J, Sampson M, Tricco AC, Moher D. Guidelines for randomized controlled trial protocol content: a systematic review. *Syst Rev* 2012;1:43.
14 Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. *Ann Intern Med* 2013. www.annals.org/article.aspx?doi=10.7326/0003-4819-158-3-201302050-00583.
15 Tetzlaff JM, Moher D, Chan A-W. Developing a guideline for reporting clinical trial protocols: Delphi consensus survey. *Trials* 2012;13:176.
16 Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.
17 Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
18 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009;62:e1-34.
19 Warner Chilcott. A comparison of once a day dose compared to 2 doses/day. http://clinicaltrials.gov/show/NCT00505778.
20 Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002;25:38-64.
21 Shaw L, Price C, McLure S, Howel D, McColl E, Ford GA. Paramedic Initiated Lisinopril For Acute Stroke Treatment (PIL-FAST): study protocol for a pilot randomised controlled trial [protocol]. *Trials* 2011;12:152.
22 Sim I, Chan A-W, Gülmezoglu AM, Evans T, Pang T. Clinical trial registration: transparency is the watchword. *Lancet* 2006;367:1631-3.
23 Dickersin K, Rennie D. Registering clinical trials. *JAMA* 2003;290:516-23.
24 Krleža-Jerić K, Chan A-W, Dickersin K, Sim I, Grimshaw J, Gluud C for the Ottawa Group. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *BMJ* 2005;330:956-8.
25 DeAngelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA* 2004;292:1363-4.
26 Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977-84.
27 Bourgeois FT, Murthy S, Mandl KD. Outcome reporting among drug trials registered in ClinicalTrials.gov. *Ann Intern Med* 2010;153:158-66.
28 You B, Gan HK, Pond G, Chen EX. Consistency in the analysis and reporting of primary end points in oncology randomized controlled trials from registration to publication: a systematic review. *J Clin Oncol* 2012;30:210-6.
29 United States Congress. Food and Drug Administration Amendments Act of 2007, Title VIII, Section 801. Expanded clinical trial registry data bank. 2007. www.govtrack.us/congress/billtext.xpd?bill=h110-3580.
30 European Commission. Communication from the Commission regarding the guideline on the data fields contained in the clinical trials database provided for in Article 11 of Directive 2001/20/EC to be included in the database on medicinal products provided for in Article 57 of Regulation (EC) No 726/2004 (2008/C 168/02). *Official Journal of the European Union* 2008;51:3-4.
31 Laine C, Horton R, DeAngelis CD, Drazen JM, Frizelle FA, Godlee F, et al. Clinical trial registration. *BMJ* 2007;334:1177-8.
32 Bernhard Nocht Institute for Tropical Medicine. Probiotic Saccharomyces boulardii for the prevention of antibiotic-associated diarrhoea (SacBo). http://clinicaltrials.gov/ct2/show/NCT01143272.
33 Dalessandro M, Hirman J. Protocol SB-275833/030—Studies 030A and 030B: two identical double-blind, double-dummy, multicenter, comparative phase III studies of the safety and efficacy of topical 1% SB-275833, applied twice daily, versus oral Cephalexin, 500 mg in adults, or 12.5 mg/kg (250 mg/5 ml) in children, twice daily, in the treatment of uncomplicated secondarily infected traumatic lesions [protocol]. Version 5 (July 25, 2005). www.spirit-statement.org/wp-content/uploads/2012/12/Protocol-SB-275833.pdf.
34 Effect of tranexamic acid on coagulation in a sample of participants in the WOMAN trial: WOMAN-ETAC study [protocol]. Version 1 (August 3, 2011). www.thewomantrial.lshtm.ac.uk/Images/WOMAN_ETACprotocol.pdf.
35 Chan A-W, Krleža-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
36 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
37 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
38 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454-65.
39 Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry* 2006;163:185-94.
40 Djulbegovic B, Cantor A, Clarke M. The importance of preservation of the ethical principle of equipoise in the design of clinical trials: relative impact of the methodological quality domains on the treatment effect in randomized controlled trials. *Account Res* 2003;10:301-15.
41 Etter J-F, Burri M, Stapleton J. The impact of pharmaceutical company funding on results of randomized trials of nicotine replacement therapy for smoking cessation: a meta-analysis. *Addiction* 2007;102:815-22.
42 Golder S, Loke YK. Is there evidence for biased reporting of published adverse effects data in pharmaceutical industry-funded studies? *Br J Clin Pharmacol* 2008;66:767-73.
43 Min Y-I, Unalp-Arida A, Scherer R, Dickersin K. Assessment of equipoise using a cohort of randomized controlled trials [abstract]. International congress on peer review and biomedical publication, Chicago, IL, 16-18 September, 2005.
44 Yaphe J, Edman R, Knishkowy B, Herman J. The association between funding by commercial interests and study outcome in randomized controlled drug trials. *Fam Pract* 2001;18:565-8.
45 Ahmer S, Arya P, Anderson D, Faruqui R. Conflict of interest in psychiatry. *Psychiatr Bull* 2005;29:302-4.
46 The Danish National Committee on Biomedical Research Ethics. Guidelines about notification etc. of a biomedical research project to the committee system on biomedical research ethics, No 9154, 5 May 2011. 2011. www.cvk.sum.dk/English/guidelinesaboutnotification.aspx.
47 Lester RT, Mills EJ, Kariri A, Ritvo P, Chung M, Jack W, et al. The HAART cell phone adherence trial (WelTel Kenya1): a randomized controlled trial protocol [protocol]. *Trials* 2009;10:87.
48 Rennie D, Yank V, Emanuel L. When authorship fails. A proposal to make contributors accountable. *JAMA* 1997;278:579-85.
49 Trials. Instructions for authors— study protocols. 2012. www.trialsjournal.com/authors/instructions/studyprotocol#formatting-contributions.
50 Williams H. Bullous Pemphigoid Steroids and Tetracyclines (BLISTER) Study. A randomised controlled trial to compare the safety and effectiveness of doxycycline (200 mg/day) with prednisolone (0.5 mg/kg/day) for initial treatment of bullous pemphigoid [protocol]. Version 4.0 (July 20, 2011). www.spirit-statement.org/wp-content/uploads/2012/12/Blister-Protocol-v4-20July2011.pdf .

51 Gertel A, Block P, Gawrylewski H-M, Raymond S, Quinn T, Muhlbradt E. CDISC Clinical research glossary. Version 8.0. 2009. www.cdisc.org/stuff/contentmgr/files/0/be650811feb46f381f0af41ca40ade2e/misc/cdisc_2009_glossary.pdf.
52 World Health Organization. Operational guidelines for ethics committees that review biomedical research. 2000. www.who.int/tdr/publications/documents/ethics.pdf.
53 World Health Organization. Handbook for good clinical research practice (GCP): Guidance for implementation. 2002. http://apps.who.int/prequal/info_general/documents/GCP/gcp1.pdf.
54 Pierce MA, Hess EP, Kline JA, Shah ND, Breslin M, Branda ME, et al. The Chest Pain Choice trial: a pilot randomized trial of a decision aid for patients with chest pain in the emergency department [protocol]. *Trials* 2010;11:57.
55 Vlad SC, LaValley MP, McAlindon TE, Felson DT. Glucosamine for pain in osteoarthritis: why do trial results differ? *Arthritis Rheum* 2007;56:2267-77.
56 Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *BMJ* 2002;325:249.
57 Liss H. Publication bias in the pulmonary/allergy literature: effect of pharmaceutical company sponsorship. *Isr Med Assoc J* 2006;8:451-4.
58 Montgomery JH, Byerly M, Carmody T, Li B, Miller DR, Varghese F, et al. An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia. *Control Clin Trials* 2004;25:598-612.
59 Perlis RH, Perlis CS, Wu Y, Hwang C, Joseph M, Nierenberg AA. Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *Am J Psychiatry* 2005;162:1957-60.
60 Jagsi R, Sheets N, Jankovic A, Motomura AR, Amarnath S, Ubel PA, et al. Frequency, nature, effects, and correlates of conflicts of interest in published clinical cancer research. *Cancer* 2009;115:2783-91.
61 Mello MM, Clarridge BR, Studdert DM. Academic medical centers' standards for clinical-trial agreements with industry. *N Engl J Med* 2005;352:2202-10.
62 European Vasculitis Study Group (EUVAS). RITUXVAS Clinical Trial Protocol: An international, randomised, open label trial comparing a rituximab based regimen with a standard cyclophosphamide/azathioprine regimen in the treatment of active, 'generalised' ANCA associated vasculitis [protocol]. Version 1b (November 15, 2005). www.vasculitis.nl/media/documents/rituxvas.pdf.
63 Delgado-Rodriguez M, Ruiz-Canela M, De Irala-Estevez J, Llorca J, Martinez-Gonzalez MA. Participation of epidemiologists and/or biostatisticians and methodological quality of published controlled clinical trials. *J Epidemiol Community Health* 2001;55:569-72.
64 Llorca J, Martinez-Sanz F, Prieto-Salceda D, Fariñas-Alvarez C, Chinchon MV, Quinones D, et al. Quality of controlled clinical trials on glaucoma and intraocular high pressure. *J Glaucoma* 2005;14:190-5.
65 CRASH2 Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage. A large randomised placebo controlled trial among trauma patients with or at risk of significant haemorrhage, of the effects of antifibrinolytic treatment on death and transfusion requirement [protocol]. Version 3 (July 2, 2005). www.crash2.lshtm.ac.uk/.
66 Clarke M. Doing new research? Don't forget the old. *PLoS Med* 2004;1:e35.
67 Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, et al. Factors that limit the quality, number and progress of randomised controlled trials. *Health Technol Assess* 1999;3:1-143.
68 Centre for Reviews and Dissemination. Systematic review of barriers, modifiers and benefits involved in participation in cancer trials. CRD Report 31. York: University of York, 2006.
69 Tournoux C, Katsahian S, Chevret S, Levy V. Factors influencing inclusion of patients with malignancies in clinical trials. *Cancer* 2006;106:258-70.
70 Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet* 2010;376:20-1.
71 Canadian Institutes of Health Research. RCT evaluation criteria and headings. 2010. www.cihr.ca/e/39187.html.
72 National Institute for Health Research. Efficacy and mechanism evaluation program. Important information & guidance notes—preliminary application. 2012. www.eme.ac.uk/funding/Researcher-led.asp.
73 Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004;364:2021-9.
74 Puhan MA, Vollenweider D, Steurer J, Bossuyt PM, ter Riet G. Where is the supporting evidence for treating mild to moderate chronic obstructive pulmonary disease exacerbations with antibiotics? A systematic review. *BMC Med* 2008;6:28.
75 Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials* 2005;2:218-29.
76 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327:248-54.
77 Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Ann Intern Med* 2011;154:50-5.
78 Goudie AC, Sutton AJ, Jones DR, Donald A. Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. *J Clin Epidemiol* 2010;63:983-91.
79 Cooper NJ, Jones DR, Sutton AJ. The use of systematic reviews when designing studies. *Clin Trials* 2005;2:260-4.
80 Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:iii-193.
81 Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ* 2012;344:d7292.
82 Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.Gov: a cross-sectional analysis. *PLoS Med* 2009;6:e1000144.
83 Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;1:MR000006.
84 Chan A-W. Out of sight but not out of mind: how to search for unpublished clinical trial evidence. *BMJ* 2012;344:d8013.
85 A phase III multi-centre, randomised, double-blind, double-dummy, comparative clinical study to assess the safety and efficacy of a fixed-dose formulation of oral pyronaridine artesunate (180:60 mg tablet) versus chloroquine (155 mg tablet), in children and adult patients with acute Plasmodium vivax malaria [protocol]. Version 2.0 (March 5, 2007). www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0014501#s5
86 Dawson L, Zarin DA, Emanuel EJ, Friedman LM, Chaudhari B, Goodman SN. Considering usual medical care in clinical trial design. *PLoS Med* 2009;6:e1000111.
87 Van Luijn JCF, Van Loenen AC, Gribnau FWJ, Leufkens HGM. Choice of comparator in active control trials of new drugs. *Ann Pharmacother* 2008;42:1605-12.
88 Johansen HK, Gøtzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. *JAMA* 1999;282:1752-9.
89 Stang A, Hense H-W, Jöckel K-H, Turner EH, Tramèr MR. Is it always unethical to use a placebo in a clinical trial? *PLoS Med* 2005;2:e72.
90 Emanuel EJ, Miller FG. The ethics of placebo-controlled trials—A middle ground. *N Engl J Med* 2001;345:915-9.
91 Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. *J Clin Epidemiol* 1999;52:1143-56.
92 Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. *Lancet Oncol* 2006;7:141-8.
93 Rochon PA, Gurwitz JH, Simms RW. A study of manufacturer supported trials of non-steroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Int Med* 1994;9:157-63.
94 Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychother Psychosom* 2009;78:172-81.
95 Sneed JR, Rutherford BR, Rindskopf D, Lane DT, Sackeim HA, Roose SP. Design makes a difference: a meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. *Am J Geriatr Psychiatry* 2008;16:65-73.
96 Sinyor M, Levitt AJ, Cheung AH, Schaffer A, Kiss A, Dowlati Y, et al. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. *J Clin Psychiatry* 2010;71:270-9.
97 Tang J-L, Zhan S-Y, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;319:160-1.
98 A phase 3, active (Warfarin) controlled, randomized, double-blind, parallel arm study to evaluate efficacy and safety of Apixaban in preventing stroke and systemic embolism in subjects with nonvalvular atrial fibrillation (ARISTOTLE: Apixaban for Reduction In STroke and Other ThromboembolLic Events in Atrial Fibrillation) [protocol]. Version 4 (August 4, 2010). www.nejm.org/doi/full/10.1056/NEJMoa1107039.
99 Fleming TR. Clinical trials: discerning hype from substance. *Ann Intern Med* 2010;153:400-406.
100 Heger U, Voss S, Knebel P, Doerr-Harim C, Neudecker J, Schuhmacher C, et al. Prevention of abdominal wound infection (PROUD trial, DRKS00000390): study protocol for a randomized controlled trial [protocol]. *Trials* 2011;12:245.
101 Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010;340:c723.
102 Dumville JC, Hahn S, Miles JN, Torgerson DJ. The use of unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials* 2006;27:1-12.
103 Gilbody S, Bower P, Torgerson D, Richards D. Cluster randomized trials produced similar results to individually randomized trials in a meta-analysis of enhanced care for depression. *J Clin Epidemiol* 2008;61:160-8.

104 Lathyris D, Trikalinos TA, Ioannidis JPA. Evidence from crossover trials: Empirical evaluation and comparison against parallel arm trials. *Int J Epidemiol* 2007;36:422-30.
105 Khan KS, Daya S, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertil Steril* 1996;65:939-45.
106 Katz J, Finnerup NB, Dworkin RH. Clinical trial outcome in neuropathic pain: relationship to study characteristics. *Neurology* 2008;70:263-72.
107 Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA* 2006;295:1147-51.
108 Fleming TR, Odem-Davis K, Rothmann MD, Li SY. Some essential considerations in the design and conduct of non-inferiority trials. *Clin Trials* 2011;8:432-9.
109 Krysan DJ, Kemper AR. Claims of equivalence in randomized controlled trials of the treatment of bacterial meningitis in children. *Pediatr Infect Dis J* 2002;21:753-8.
110 Tinmouth JM, Steele LS, Tomlinson G, Glazier GH. Are claims of equivalency in digestive diseases trials supported by the evidence. *Gastroenterol* 2004;126:1700-10.
111 Kairalla JA, Coffey CS, Thomann MA, Muller KE. Adaptive trial designs: a review of barriers and opportunities. *Trials* 2012;13:145.
112 Dragalin V. Adaptive designs: terminology and classification. *Drug Inf J* 2006;40:425-35.
113 Project Accept Study Group. Project Accept (HPTN 043): A phase III randomized controlled trial of community mobilization, mobile testing, same-day results, and post-test support for HIV in Sub-Saharan Africa and Thailand [protocol]. Version 2.4 (April 15, 2011). www.hptn.org/research_studies/hptn043.asp.
114 Ford JG, Howerton MW, Lai GY, Gary TL, Bolen S, Gibbons MC, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. *Cancer* 2008;112:228-42.
115 Elkins JS, Khatabi T, Fung L, Rootenberg J, Johnston SC. Recruiting subjects for acute stroke trials: a meta-analysis. *Stroke* 2006;37:123-8.
116 Heo M, Papademetriou E, Meyers BS. Design characteristics that influence attrition in geriatric antidepressant trials: meta-analysis. *Int J Geriatr Psychiatry* 2009;24:990-1001.
117 Fabricatore AN, Wadden TA, Moore RH, Butryn ML, Gravallese EA, Erondu NE, et al. Attrition from randomized controlled trials of pharmacological weight loss agents: a systematic review and analysis. *Obes Rev* 2009;10:333-41.
118 Lemieux J, Goodwin PJ, Pritchard KI, Gelmon KA, Bordeleau LJ, Duchesne T, et al. Identification of cancer care and protocol characteristics associated with recruitment in breast cancer clinical trials. *J Clin Oncol* 2008;26:4458-65.
119 Jones R, Jones RO, McCowan C, Montgomery AA, Fahey T, Jones R, et al. The external validity of published randomized controlled trials in primary care. *BMC Fam Pract* 2009;10:5.
120 Sood A, Knudsen K, Sood R, Wahner-Roedler DL, Barnes SA, Bardia A, et al. Publication bias for CAM trials in the highest impact factor medicine journals is partly due to geographical bias. *J Clin Epidemiol* 2007;60:1123-6.
121 Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials* 2009;10:46.
122 Hotopf M, Lewis G, Normand C. Putting trials on trial--the costs and consequences of small trials in depression: a systematic review of methodology. *J Epidemiol Community Health* 1997;51:354-8.
123 Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness (ESCAPE) [protocol]. Version 3.0 (November 29, 1999). https://biolincc.nhlbi.nih.gov/studies/escape/?q=escape.
124 Sandercock P, Lindley R, Wardlaw J, Dennis M, Lewis S, Venables G, et al. The third international stroke trial (IST-3) of thrombolysis for acute ischaemic stroke [protocol]. *Trials* 2008;9:37.
125 Blümle A, Meerpohl JJ, Rücker G, Antes G, Schumacher M, von Elm E. Reporting of eligibility criteria of randomised trials: cohort study comparing trial protocols with subsequent articles. *BMJ* 2011;342:d1828.
126 Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials* 2009;10:9.
127 Simpson F, Sweetman EA, Doig GS. Systematic review of techniques and interventions for improving adherence to inclusion and exclusion criteria during enrolment into randomised controlled trials. *Trials* 2010;11:17.
128 Rendell JM, Merritt RK, Geddes JR. Incentives and disincentives to participation by clinicians in randomised controlled trials. *Cochrane Database Syst Rev* 2007;2:MR000021.
129 Weijer C. Characterizing the population in clinical trials: barriers, comparability, and implications for review. *Philosophy Publications.* Paper 250.1995. http://ir.lib.uwo.ca/philosophypub/250.
130 Townsley CA, Selby R, Siu LL. Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials. *J Clin Oncol* 2005;23:3112-24.
131 Uchino K, Billheimer D, Cramer SC. Entry criteria and baseline characteristics predict outcome in acute stroke trials. *Stroke* 2001;32:909-16.
132 Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 2007;297:1233-40.
133 Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epimiol* 2000;53:973-9.
134 Gandhi M, Ameli N, Bacchetti P, Sharp GB, French AL, Young M, et al. Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS* 2005;19:1885-96.
135 Montori VM, Wang YG, Alonso-Coello P, Bhagra S. Systematic evaluation of the quality of randomized controlled trials in diabetes. *Diabetes Care* 2006;29:1833-8.
136 Mitchell SL, Sullivan EA, Lipsitz LA. Exclusion of elderly subjects from clinical trials for Parkinson disease. *Arch Neurol* 1997;54:1393-8.
137 Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ* 2009;180:E47-57.
138 Blanco C, Olfson M, Goodwin RD, Ogburn E, Liebowitz MR, Nunes EV, et al. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry* 2008;69:1276-80.
139 Herland K, Akselsen JP, Skjønsberg OH, Bjermer L. How representative are clinical study patients with asthma or COPD for a larger "real life" population of patients with obstructive lung disease? *Respir Med* 2005;99:11-9.
140 Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess* 2005;9:iii-iiv.
141 Zarin DA, Young JL, West JC. Challenges to evidence-based medicine: a comparison of patients and treatments in randomized controlled trials with patients and treatments in a practice research network. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:27-35.
142 Hordijk-Trion M, Lenzen M, Wijns W, de Jaegere P, Simoons ML, Scholte op Reimer WJ, et al. Patients enrolled in coronary intervention trials are not representative of patients in clinical practice: results from the Euro Heart Survey on Coronary Revascularization. *Eur Heart J* 2006;27:671-8.
143 Kievit W, Fransen J, Oerlemans AJ, Kuper HH, van der Laar MA, de Rooij DJ, et al. The efficacy of anti-TNF in rheumatoid arthritis, a comparison between randomised controlled trials and clinical practice. *Ann Rheum Dis* 2007;66:1473-8.
144 Uijen AA, Bakx JC, Mokkink HG, van Weel C. Hypertension patients participating in trials differ in many aspects from patients treated in general practices. *J Clin Epidemiol* 2007;60:330-5.
145 Crossman DC, Morton AC, Gunn JP, Greenwood JP, Hall AS, Fox KA, et al. Investigation of the effect of Interleukin-1 receptor antagonist (IL-1ra) on markers of inflammation in non-ST elevation acute coronary syndromes (The MRC-ILA-HEART Study) [protocol]. *Trials* 2008;9:8.
146 Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ* 2008;336:1472-4.
147 Duff JM, Leather H, Walden EO, LaPlant KD, George TJ, Jr. Adequacy of published oncology randomized controlled trials to provide therapeutic details needed for clinical application. *J Natl Cancer Inst* 2010;102:702-5.
148 Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86-9.
149 Glasziou P, Chalmers I, Altman DG, Bastian H, Boutron I, Brice A, et al. Taking healthcare interventions from trial to practice. *BMJ* 2010;341:c3852.
150 Golomb BA, Erickson LC, Koperski S, Sack D, Enkin M, Howick J. What's in placebos: who knows? Analysis of randomized, controlled trials. *Ann Intern Med* 2010;153:532-5.
151 Medical Research Council Working Party on Prostate Cancer. MRC PR05. A Medical Research Council randomised trial of adjuvant sodium clodronate in patients commencing or responding to hormone therapy for metastatic prostate adenocarcinoma [protocol]. Feb 1995 version. www.ctu.mrc.ac.uk/research_areas/study_details.aspx?s=60.
152 Panel on Handling Missing Data in Clinical Trials, National Research Council. The prevention and treatment of missing data in clinical trials. Washington DC, National Academies Press, 2010.
153 Buchbinder S, Liu A, Thompson M, Mayer K. Phase II extended safety study of tenofovir disoproxil fumarate (TDF) among HIV-1 negative men [protocol]. Version 1.6 (February 16, 2007). www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0023688.
154 World Health Organization. Adherence to long-term therapies: evidence for action. 2012. www.who.int/chp/knowledge/publications/adherence_full_report.pdf.
155 Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med* 2005;353:487-97.
156 Smith D. Patient nonadherence in clinical trials: could there be a link to postmarketing patient safety? *Drug Inf J* 2012;46:27-34.
157 Robiner WN. Enhancing adherence in clinical research. *Contemp Clin Trials* 2005;26:59-77.
158 Matsui D. Strategies to measure and improve patient adherence in clinical trials. *Pharmaceut Med* 2009;23:289-97.

159 Simpson SH, Eurich DT, Majumdar SR, Padwal RS, Tsuyuki RT, Varney J, et al. A meta-analysis of the association between adherence to drug therapy and mortality. *BMJ* 2006;333:15.
160 International Conference on Harmonisation. ICH Harmonised Tripartite Guideline: Good clinical practice, consolidated guideline. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (June 1996, E6). 1996. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf.
161 Jayaraman S, Rieder MJ, Matsui DM. Compliance assessment in drug trials: has there been improvement in two decades? *Can J Clin Pharmacol* 2005;12:e251-3.
162 Sackett DL. Clinician-trialist rounds: 5. Cointervention bias--how to diagnose it in their trial and prevent it in yours. *Clin Trials* 2011;8:440-2.
163 Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. *N Engl J Med* 2011;364:852-60.
164 Bhandari M, Lochner H, Tornetta P, III. Effect of continuous versus dichotomous outcome variables on study power when sample sizes of orthopaedic randomized trials are small. *Arch Orthop Trauma Surg* 2002;122:96-8.
165 Verhagen AP, de Vet HCW, Willemsen S, Stijnen T. A meta-regression analysis shows no impact of design characteristics on outcome in trials on tension-type headaches. *J Clin Epi* 2008;61:813-8.
166 Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012;344:e1119.
167 Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429-8.
168 Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007;334:786.
169 Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ* 2005;330:596.
170 Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;289:2554-59.
171 Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ* 2010;341:c3920.
172 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;3:e3081.
173 Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: Review of publication and presentation. *PLoS Med* 2008;5:e217.
174 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
175 Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med* 2009;361:1963-71.
176 Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR. Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database Syst Rev* 2011;1:MR000031.
177 Chan A-W. Access to clinical trial data. *BMJ* 2011;342:d80.
178 Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials* 2007;8:38.
179 Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy* 2012;17:1-2.
180 Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39.
181 Booth R, Fuller B, Thompson L, McCarty D, Shoptaw S, et al. STUDY #: NIDA-CTN-0017. HIV and HCV risk reduction interventions in drug detoxification and treatment settings [protocol]. Version 4.0 (August 16, 2010). www.dtmi.duke.edu/crflibrary-demo/crf-library-1/crf-library/trials-a-e/ctn-0017.
182 Cockayne NL, Glozier N, Naismith SL, Christensen H, Neal B, Hickie IB. Internet-based treatment for older adults with depression and co-morbid cardiovascular disease: protocol for a randomised, double-blind, placebo controlled trial [protocol]. *BMC Psychiatry* 2011;11:10.
183 McMurran M, Crawford MJ, Reilly JG, McCrone P, Moran P, Williams H, et al. Psycho-education with problem solving (PEPS) therapy for adults with personality disorder: A pragmatic multi-site community-based randomised clinical trial [protocol]. *Trials* 2011;12:198.
184 van der Lee JH, Wesseling J, Tanck MW, Offringa M. Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases. *J Clin Epidemiol* 2008;61:324-30.
185 Yazici Y, Adler NM, Yazici H. Most tumour necrosis factor inhibitor trials in rheumatology are undeservedly called 'efficacy and safety' trials: a survey of power considerations. *Rheumatol* 2008;47:1054-7.
186 Hernández AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J* 2006;151:257-64.
187 Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541-6.
188 Raju TN, Langenberg P, Sen A, Aldana O. How much 'better' is good enough? The magnitude of treatment effect in clinical trials. *Am J Dis Child* 1992;146:407-11.
189 Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.
190 Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol* 2003;56:717-20.
191 Proschan MA. Sample size re-estimation in clinical trials. *Biom J* 2009;51:348-57.
192 Julious SA, Campbell MJ, Altman DG. Estimating sample sizes for continuous, binary, and ordinal outcomes in paired comparisons: practical hints. *J Biopharm Stat* 1999;9:241-51.
193 Campbell MK, Elbourne DR, Altman DG, CONSORT group. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328:702-8.
194 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA* 2006;295:1152-60.
195 Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. *Am J Pub Health* 2008;98:1418-24.
196 Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: Systematic review of recent trials. *BMJ* 2008;336:876-80.
197 Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials* 2004;1:80-90.
198 Murray DM, Pals SL, Blitstein JL, Alfano CM, Lehman J. Design and analysis of group-randomized trials in cancer: A review of current practices. *J Natl Cancer Inst* 2008;100:483-91.
199 Freiman JA, Chalmers TC, Smith H, Jr., Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978;299:690-4.
200 Bailey CS, Fisher CG, Dvorak MF. Type II error in the spine surgical literature. *Spine* 2004;29:1146-9.
201 Lochner HV, Bhandari M, Tornetta P, III. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am* 2001;83-A:1650-5.
202 Enwere G. A review of the quality of randomized clinical trials of adjunctive therapy for the treatment of cerebral malaria. *Trop Med Int Health* 2005;10:1171-5.
203 Breau RH, Carnat TA, Gaboury I. Inadequate statistical power of negative clinical trials in urological literature. *J Urol* 2006;176:263-6.
204 Keen HI, Pile K, Hill CL. The prevalence of underpowered randomized clinical trials in rheumatology. *J Rheumatol* 2005;32:2083-8.
205 Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: are they done correctly? *Surgery* 2003;134:275-9.
206 Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg* 2001;136:796-800.
207 Murray GD. Research governance must focus on research training. *BMJ* 2001;322:1461-2.
208 Asthma Clinical Research Network. Beta Adrenergic Response by Genotype (BARGE) study protocol: a study to compare the effects of regularly scheduled use of inhaled albuterol in patients with mild to moderate asthma who are members of two distinct haplotypes expressed at the β2 -adrenergic receptor [protocol]. Version 5.4 (September 23, 1999). https://biolincc.nhlbi.nih.gov/studies/barge/?q=barge.
209 Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knights R, et al. Recruitment to randomised trials: Strategies for trial enrolment and participation study. The STEPS study. *Health Technol Assess* 2007;11:iii-72.
210 Wise P, Drury M. Pharmaceutical trials in general practice: the first 100 protocols. An audit by the clinical research ethics committee of the Royal College of General Practitioners. *BMJ* 1996;313:1245-8.
211 Pich J, Carné X, Arnaiz JA, Gómez B, Trilla A, Rodés J. Role of a research ethics committee in follow-up and publication of results. *Lancet* 2003;361:1015-6.
212 Decullier E, Lhéritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ* 2005;331:19.
213 Dal-Ré R, Ortega R, Espada J. [Efficiency of investigators in recruitment of patients for clinical trials: apropos of a multinational study]. *Med Clin (Barc)* 1998;110:521-3.

214 McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials* 2006;7:9.
215 Charlson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *BMJ* 1984;289:1281-4.
216 Caldwell PH, Hamilton S, Tan A, Craig JC. Strategies for increasing recruitment to randomised controlled trials: systematic review. *PLoS Med* 2010;7:e1000368.
217 Treweek S, Pitkethly M, Cook J, Kjeldstrøm M, Taskila T, Johansen M, et al. Strategies to improve recruitment to randomised controlled trials. *Cochrane Database Syst Rev* 2010;4:MR000013.
218 Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. *Surgery* 2006;139:469-83.
219 Lai GY, Gary TL, Tilburt J, Bolen S, Baffi C, Wilson RF, et al. Effectiveness of strategies to recruit underrepresented populations into cancer clinical trials. *Clin Trials* 2006;3:133-41.
220 UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. *J Gen Intern Med* 2007;22:852-63.
221 Miller NL, Markowitz JC, Kocsis JH, Leon AC, Brisco ST, Garno JL. Cost effectiveness of screening for clinical trials by research assistants versus senior investigators. *J Psychiatr Res* 1999;33:81-5.
222 Tworoger SS, Yasui Y, Ulrich CM, Nakamura H, LaCroix K, Johnston R, et al. Mailing strategies and recruitment into an intervention trial of the exercise effect on breast cancer biomarkers. *Cancer Epidemiol Biomarkers Prev* 2002;11:73-7.
223 Schroy P.C. 3rd, Glick JT, Robinson P, Lydotes MA, Heeren TC, Prout M, et al. A cost-effectiveness analysis of subject recruitment strategies in the HIPAA era: results from a colorectal cancer screening adherence trial. *Clin Trials* 2009;6:597-609.
224 Harvey LA, Dunlop SA, Churilov L, Hsueh Y-SA, Galea MP. Early intensive hand rehabilitation after spinal cord injury ("hands on"): a protocol for a randomised controlled trial [protocol]. *Trials* 2011;12:14.
225 Schulz KF, Grimes DA. The Lancet handbook of essential concepts in clinical research. Elsevier, 2006.
226 Greenland S. Randomization, statistics, and causal inference. *Epidemiol* 1990;1:421-9.
227 Armitage P. The role of randomization in clinical trials. *Stat Med* 1982;1:345-52.
228 Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011;4:MR000012.
229 Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
230 McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug Inf J* 2003;37:293-308.
231 Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359:515-9.
232 Altman DG, Bland JM. How to randomise. *BMJ* 1999;319:703-4.
233 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
234 Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol* 1999;52:19-26.
235 Han B, Enas NH, McEntegart D. Randomization by minimization for unbalanced treatment allocation. *Stat Med* 2009;28:3329-46.
236 Altman DG. Practical statistics for medical research. Chapman and Hall/CRC, 1991.
237 Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* 1998;317:362-3.
238 Berger VW. Varying the block size does not conceal the allocation. *J Crit Care* 2006;21:229-30.
239 Berger VW. Minimization, by its nature, precludes allocation concealment, and invites selection bias. *Contemp Clin Trials* 2010;31:406.
240 Abbott JH, Robertson MC, McKenzie JE, Baxter GD, Theis J-C, Campbell AJ, et al. Exercise therapy, manual therapy, or both, for osteoarthritis of the hip or knee: a factorial randomised controlled trial protocol [protocol]. *Trials* 2009;10:11.
241 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002;359:614-618.
242 Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med* 1987;6:315-28.
243 Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;272:125-8.
244 Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. *J Clin Epidemiol* 2011;64:1070-5.
245 Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007;2:MR000012.
246 Klingberg S, Wittorf A, Meisner C, Wölwer W, Wiedemann G, Herrlich J, et al. Cognitive behavioural therapy versus supportive therapy for persistent positive symptoms in psychotic disorders: The POSITIVE study, a multicenter, prospective, single-blind, randomised controlled clinical trial [protocol]. *Trials* 2010;11:123.
247 Dalum HS, Korsbek L, Mikkelsen JH, Thomsen K, Kistrup K, Olander M, et al. Illness management and recovery (IMR) in Danish community mental health centres [protocol]. *Trials* 2011;12:195.
248 Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Syst Rev* 2010;1:CD003974.
249 Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005;34:79-87.
250 Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Bürgi E, Scherer M, et al. The effects of excluding patients from the analysis in randomised controlled trials: meta-epidemiological study. *BMJ* 2009;339:b3244.
251 Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254-59.
252 Ballintine EJ. Randomized controlled clinical trial. National Eye Institute workshop for ophthalmologists. Objective measurements and the double-masked procedure. *Am J Ophthalmol* 1975;79:763-7.
253 Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials* 1996;17:285-90.
254 Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess* 2005;9:1-238.
255 Meinert CL. Masked monitoring in clinical trials—blind stupidity? *N Engl J Med* 1998;338:1381-2.
256 Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacological treatments: a systematic review. *PLoS Med* 2006;3:e425.
257 Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007;4:e61.
258 Lieverse R, Nielen MM, Veltman DJ, Uitdehaag BM, van Someren EJ, Smit JH, et al. Bright light in elderly subjects with nonseasonal major depressive disorder: a double blind randomised clinical trial using early morning bright blue light comparing dim red light treatment. *Trials* 2008;9:48.
259 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000-3.
260 Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials* 2006;3:360-5.
261 Hróbjartsson A, Boutron I. Blinding in randomized clinical trials: imposed impartiality. *Clin Pharmacol Ther* 2011;90:732-6.
262 Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004;328:432.
263 Sackett DL. Clinician-trialist rounds: 6. Testing for blindness at the end of your trial is a mug's game. *Clin Trials* 2011;8:674-6.
264 Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. *Lancet* 2010;375:1144-6.
265 A randomized, double blind, placebo controlled, parallel group trial for assessing the clinical benefit of Dronedarone 400mg BID on top of standard therapy in patients with permanent atrial fibrillation and additional risk factors. Permanent Atrial fibriLLAtion outcome Study using Dronedarone on top of standard therapy (PALLAS) [protocol]. Version 1 (February 26, 2010). www.nejm.org/doi/full/10.1056/NEJMoa1109867.
266 Campbell NL, Khan BA, Farber M, Campbell T, Perkins AJ, Hui SL, et al. Improving delirium care in the intensive care unit: the design of a pragmatic study [protocol]. *Trials* 2011;12:139.
267 FSGS - Clinical trial [protocol]. Version 3c (June 20, 2005). https://clinicalresearch.ccf.org/fsgs/docs/index_docs.html.
268 Lane SJ, Heddle NM, Arnold E, Walker I. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. *BMC Med Inform Decis Mak* 2006;6:23.
269 Bent S, Padula A, Avins AL. Brief communication: Better ways to question patients about adverse medical events: a randomized, controlled trial. *Ann Intern Med* 2006;144:257-61.
270 Dale O, Hagen KB. Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data. *J Clin Epidemiol* 2007;60:8-17.
271 Litchfield J, Freeman J, Schou H, Elsley M, Fuller R, Chubb B. Is the future for clinical trials internet-based? A cluster randomised clinical trial. *Clin Trials* 2005;2:72-9.
272 Bedard M, Molloy DW, Standish T, Guyatt GH, D'Souza J, Mondadori C, et al. Clinical trials in cognitively impaired older adults: home versus clinic assessments. *J Am Geriatr Soc* 1995;43:1127-30.
273 Jasperse DM, Ahmed SW. The Mid-Atlantic Oncology Program's comparison of two data collection methods. *Control Clin Trials* 1989;10:282-9.
274 Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009;101:1624-32.

275 Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. *Rheumatology (Oxford)* 2004;43:704-11.
276 Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol* 2004;22:3485-90.
277 Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. *PLoS One* 2011;6:e25348.
278 Kryworuchko J, Stacey D, Bennett C, Graham ID. Appraisal of primary outcome measures used in trials of patient decision support. *Patient Educ Couns* 2008;73:497-503.
279 Roberts L, Counsell C. Assessment of clinical outcomes in acute stroke trials. *Stroke* 1998;29:986-91.
280 Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249-52.
281 Williams GW. The other side of clinical trial monitoring; assuring data quality and procedural adherence. *Clin Trials* 2006;3:530-7.
282 Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP. Data quality assurance, monitoring, and reporting. *Control Clin Trials* 1995;16:104S-36S.
283 Meyerson LJ, Wiens BL, LaVange LM, Koutsoukos AD. Quality control of oncology clinical trials. *Hematol Oncol Clin North Am* 2000;14:953-71.
284 Fong DYT. Data management and quality assurance. *Drug Inf J* 2001;35:839-44.
285 Knatterud GL, Rockhold FW, George SL, Barton FB, Davis CE, Fairweather WR, et al. Guidelines for quality assurance in multicenter trials: a position paper. *Control Clin Trials* 1998;19:477-93.
286 Prevention Study Group. HEALTHY primary prevention trial protocol [protocol]. Version 1.4 (July 14, 2008). www.healthystudy.org/.
287 HIV Prevention Trials Network and the International Maternal Pediatric and Adolescent AIDS Clinical Trials Network. HPTN 046: A phase III trial to determine the efficacy and safety of an extended regimen of nevirapine in infants born to HIV-infected women to prevent vertical HIV transmission during breastfeeding [protocol]. Version 3.0 (September 26, 2007). www.hptn.org/research_studies/hptn046.asp.
288 Ioannidis JP, Bassett R, Hughes MD, Volberding PA, Sacks HS, Lau J. Predictors and impact of patients lost to follow-up in a long-term randomized trial of immediate versus deferred antiretroviral treatment. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997;16:22-30.
289 Ford ME, Havstad S, Vernon SW, Davis SD, Kroll D, Lamerato L, et al. Enhancing adherence among older African American men enrolled in a longitudinal cancer screening trial. *Gerontologist* 2006;46:545-50.
290 Couper MP, Peytchev A, Strecher VJ, Rothert K, Anderson J. Following up nonrespondents to an online weight management intervention: Randomized trial comparing mail versus telephone. *J Med Internet Res* 2007;9:e16.
291 Renfroe EG, Heywood G, Foreman L, Schron E, Powell J, Baessler C, et al. The end-of-study patient survey: methods influencing response rate in the AVID Trial. *Control Clin Trials* 2002;23:521-33.
292 Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. *J Clin Epi* 2007;60:757-65.
293 Fleming TR. Addressing missing data in clinical trials. *Ann Intern Med* 2011;154:113-7.
294 Liu M, Wei L, Zhang J. Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharm Stat* 2006;5:7-18.
295 Wahlbeck K, Tuunainen A, Ahokas A, Leucht S. Dropout rates in randomised antipsychotic drug trials. *Psychopharmacology (Berl)* 2001;155:230-33.
296 Kawado M, Hinotsu S, Matsuyama Y, Yamaguchi T, Hashimoto S, Ohashi Y. A comparison of error detection rates between the reading aloud method and the double data entry method. *Control Clin Trials* 2003;24:560-9.
297 Day S, Fayers P, Harvey D. Double data entry: what value, what price? *Control Clin Trials* 1998;19:15-24.
298 Reynolds-Haertle RA, McBride R. Single vs. double data entry in CAST. *Control Clin Trials* 1992;13:487-94.
299 Gibson D, Harvey AJ, Everett V, Parmar MK. Is double data entry necessary? The CHART trials. CHART Steering Committee. Continuous, hyperfractionated, accelerated radiotherapy. *Control Clin Trials* 1994;15:482-8.
300 Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz KF, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
301 Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. *Lancet* 2005;365:1591-5.
302 Tendal B, Nüesch E, Higgins JP, Jüni P, Gøtzsche PC. Multiplicity of data in trial reports and the reliability of meta-analyses: empirical study. *BMJ* 2011;343:d4829.
303 Flow Investigators. Fluid lavage of open wounds (FLOW): design and rationale for a large, multicenter collaborative 2 x 3 factorial trial of irrigating pressures and solutions in patients with open fractures [protocol]. *BMC Musculoskelet Disord* 2010;11:85.
304 Resuscitation Outcomes Consortium Prehospital Resuscitation using an IMpedance valve and Early vs Delayed analysis (ROC PRIMED) Trial. A factorial design of an active impedence threshold valve versus sham valve and analyze later versus analyze early [protocol]. Dec 2006 version. www.nejm.org/doi/full/10.1056/NEJMoa1010821.
305 Boonacker CW, Hoes AW, van Liere-Visser K, Schilder AG, Rovers MM. A comparison of subgroup analyses in grant applications and publications. *Am J Epidemiol* 2011;174:219-25.
306 Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 2005;365:1657-61.
307 Hirji KF, Fagerland MW. Outcome based subgroup analysis: a neglected concern. *Trials* 2009;10:33.
308 Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117.
309 Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176-86.
310 Yu L-M, Chan A-W, Hopewell S, Deeks JJ, Altman DG. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review. *Trials* 2010;11:59.
311 Chen X, Liu M, Zhang J. A note on postrandomization adjustment of covariates. *Drug Inf J* 2005;39:373-83.
312 Rochon J. Issues in adjusting for covariates arising postrandomization in clinical trials. *Drug Inf J* 1999;33:1219-28.
313 Mohr JP, Moskowitz A, Ascheim D, Gelijns A, Parides M, et al. A Randomized multicenter clinical trial of unruptured brain AVMs (ARUBA): clinical protocol [protocol]. Version 3.0 (October 16, 2008). http://research.ncl.ac.uk/nctu/ARUBA.html.
314 Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ* 2010;340:c2697.
315 Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325:652-4.
316 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.
317 Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ* 2012;344:e2809.
318 Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials* 2004;1:368-76.
319 Fielding S, Fayers P, Ramsay CR. Analysing randomised controlled trials with missing data: Choice of approach affects conclusions. *Contemp Clin Trials* 2012;33:461-9.
320 Streiner DL. Missing data and the trouble with LOCF. *Evid Based Ment Health* 2008;11:3-5.
321 Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.
322 Groenwold RH, Donders AR, Roes KC, Harrell FE, Jr., Moons KG. Dealing with missing outcome data in randomized trials and observational studies. *Am J Epidemiol* 2012;175:210-7.
323 Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med* 2009;6:e1000065.
324 Berger VW. Conservative handling of missing data. *Contemp Clin Trials* 2012;33:460.
325 Azuara-Blanco A, Burr JM, Cochran C, Ramsay C, Vale L, Foster P, et al. The effectiveness of early lens extraction with intraocular lens implantation for the treatment of primary angle-closure glaucoma (EAGLE): study protocol for a randomized controlled trial [protocol]. *Trials* 2011;12:133.
326 Sydes MR, Altman DG, Babiker AB, Parmar MK, Spiegelhalter DJ, DAMOCLES Group. Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study. *Clin Trials* 2004;1:48-59.
327 Floriani I, Rotmensz N, Albertazzi E, Torri V, De Rosa M, Tomino C, et al. Approaches to interim analysis of cancer randomised clinical trials with time to event endpoints: a survey from the Italian National Monitoring Centre for Clinical Trials. *Trials* 2008;9:46.
328 Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *JAMA* 2012;307:1838-47.
329 Ellenberg SS. Independent data monitoring committees: rationale, operations and controversies. *Stat Med* 2001;20:2573-2583.
330 Ellenberg SS, Fleming TR, DeMets DL. Data monitoring committees in clinical trials: a practical perspective. 6th ed. Wiley, 2002.
331 DAMOCLES study group, NHS Health Technology Assessment Programme. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet* 2005;365:711-22.
332 Bakker OJ, van Santvoort HC, van Brunschot S, Ali UA, Besselink MG, et al. Pancreatitis, very early compared with normal start of enteral feeding (PYTHON trial): design and rationale of a randomised controlled multicenter trial [protocol]. *Trials* 2011;12:73.
333 DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. *Lancet* 1999;354:1983-8.

334 Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med* 1985;4:521-6.
335 Pocock SJ. When to stop a clinical trial. *BMJ* 1992;305:235-40.
336 Aronson JK, Ferner RE. Clarification of terminology in drug safety. *Drug Saf* 2005;28:851-70.
337 Myers MG, Cairns JA, Singer J. The consent form as a possible cause of side effects. *Clin Pharmacol Ther* 1987;42:250-3.
338 Wallin J, Sjövall J. Detection of adverse drug reactions in a clinical trial using two types of questioning. *Clin Ther* 1981;3:450-2.
339 Gøtzsche PC. Non-steroidal anti-inflammatory drugs. *BMJ* 2000;320:1058-61.
340 Curfman GD, Morrissey S, Drazen JM. Expression of concern reaffirmed. *N Engl J Med* 2006;354:1193.
341 Wright JM, Perry TL, Bassett KL, Chambers GK. Reporting of 6-month vs 12-month data in a clinical trial of celecoxib. *JAMA* 2001;286:2398-400.
342 Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials* 2009;6:430-40.
343 Sherman RB, Woodcock J, Norden J, Grandinetti C, Temple RJ. New FDA regulation to improve safety reporting in clinical trials. *N Engl J Med* 2011;365:3-5.
344 Ruiz-Canela M, Martinez-González MA, Gómez-Gracia E, Fernández-Crehuet J. Informed consent and approval by institutional review boards in published reports on clinical trials. *N Engl J Med* 1999;340:1114-5.
345 Breast Cancer International Research Group. BCIRG 006: Multicenter phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel (AC-->T) with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab (Herceptin®) (AC-->TH) and with docetaxel, carboplatin and trastuzumab (TCH) in the adjuvant treatment of node positive and high risk node negative patients with operable breast cancer containing the HER2 alteration [protocol]. Version 5 www.nejm.org/doi/full/10.1056/NEJMoa0910383.
346 Getz KA, Zuckerman R, Cropp AB, Hindle AL, Krauss R, Kaitin KI. Measuring the incidence, causes, and repercussions of protocol amendments. *Drug Inf J* 2011;45:265-75.
347 Decullier E, Lhéritier V, Chapuis F. The activity of French research ethics committees and characteristics of biomedical research protocols involving humans: a retrospective cohort study. *BMC Med Ethics* 2005;6:e9.
348 Lösch C, Neuhäuser M. The statistical analysis of a clinical trial when a protocol amendment changed the inclusion criteria. *BMC Med Res Methodol* 2008;8:16.
349 US Food and Drug Administration. Code of federal regulations. Title 21, Vol 5. 21CFR312.30. 2011.
350 European Commission. Communication from the Commission—Detailed guidance on the request to the competent authorities for authorisation of a clinical trial on a medicinal product for human use, the notification of substantial amendments and the declaration of the end of the trial (CT-1) (2010/C 82/01). *Off J European Union* 2010;53.
351 Bond J, Wilson J, Eccles M, Vanoli A, Steen N, Clarke R, et al. Protocol for north of England and Scotland study of tonsillectomy and adeno-tonsillectomy in children (NESSTAC). A pragmatic randomised controlled trial comparing surgical intervention with conventional medical treatment in children with recurrent sore throats [protocol]. *BMC Ear,Nose Throat Disord* 2006;6:13.
352 Williams CJ, Zwitter M. Informed consent in European multicentre randomised clinical trials - Are patients really informed? *Eur J Cancer* 1994;30:907-10.
353 Ryan RE, Prictor MJ, McLaughlin KJ, Hill SJ. Audio-visual presentation of information for informed consent for participation in clinical trials. *Cochrane Database Syst Rev* 2008;1:CD003717.
354 Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *JAMA* 2004;292:1593-601.
355 Cohn E, Larson E. Improving participant comprehension in the informed consent process. *J Nurs Scholarsh* 2007;39:273-80.
356 Wendler DS. Assent in paediatric research: theoretical and practical considerations. *J Med Ethic* 2006;32:229.
357 McRae AD, Weijer C, Binik A, Grimshaw JM, Boruch R, Brehaut JC, et al. When is informed consent required in cluster randomized trials in health research? *Trials* 2011;12:202.
358 Beskow LM, Friedman JY, Hardy NC, Lin L, Weinfurt KP. Developing a simplified consent form for biobanking. *PLoS One* 2010;5:e13302.
359 HIV Prevention Trials Network. HPTN 037: A phase III randomized study to evaluate the efficacy of a network-oriented peer educator intervention for the prevention of HIV transmission among injection drug users and their network members [protocol]. Version 2.0 (October 23, 2003). www.hptn.org/research_studies/hptn037.asp.
360 World Association of Medical Editors Editorial Policy and Publication Ethics Committees. Conflict of interest in peer-reviewed medical journals. 2009. www.wame.org/conflict-of-interest-in-peer-reviewed-medical-journals.
361 Rochon PA, Hoey J, Chan A-W, Ferris LE, Lexchin J, Kalkar SR, et al. Financial conflicts of interest checklist 2010 for clinical research studies. *Open Med* 2010;4:e69-91.
362 Drazen JM, de Leeuw PW, Laine C, Mulrow C, DeAngelis CD, Frizelle FA, et al. Towards more uniform conflict disclosures: the updated ICMJE conflict of interest reporting form. *BMJ* 2010;340:c3239.
363 World Medical Association. WMA statement on conflict of interest. 2012. www.wma.net/en/30publications/10policies/i3/.
364 Lundh A, Krogsbøll LT, Gøtzsche PC. Access to data in industry-sponsored trials. *Lancet* 2011;378:1995-6.
365 Microbicide Trials Network. MTN-003: Phase 2B safety and effectiveness study of tenofovir 1% gel, tenofovir disproxil fumarate tablet and emtricitabine/tenofovir disoproxil fumarate tablet for the prevention of HIV infection in women [protocol]. Version 2.0 (December 31, 2010). www.mtnstopshiv.org/news/studies/mtn003.
366 Richardson HS, Belsky L. The ancillary-care responsibilities of medical researchers. *Hastings Center Report* 2004;34:25-33.
367 Belsky L, Richardson HS. Medical researchers' ancillary clinical care responsibilities. *BMJ* 2004;328:1494-6.
368 Sofaer N, Strech D. Reasons why post-trial access to trial drugs should, or need not be ensured to research participants: A systematic review. *Public Health Ethics* 2011;4:160-84.
369 Participants in the 2006 Georgetown University Workshop on the Ancillary-Care Obligations of Medical Researchers Working in Developing Countries. The ancillary-care obligations of medical researchers working in developing countries. *PLoS Med* 2008;5:e90.
370 Beta-Blocker Evaluation of Survival Trial (BEST) Protocol [protocol]. Version 1 (June 22, 1999). https://biolincc.nhlbi.nih.gov/studies/best/.
371 Mann H. Research ethics committees and public dissemination of clinical trial results. *Lancet* 2002;360:406-8.
372 Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. *Trials* 2011;12:249.
373 Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet* 2004;363:1341-5.
374 Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *Int J Cardiol* 1993;40:161-6.
375 McGauran N, Wieseler B, Kreis J, Schüler YB, Kölsch H, Kaiser T. Reporting bias in medical research - a narrative review. *Trials* 2010;11:37.
376 Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ* 2012;344:d7202.
377 Doshi P, Jones M, Jefferson T. Rethinking credible evidence synthesis. *BMJ* 2012;344:d7898.
378 Emerson GB, Warme WJ, Wolf FM, Heckman JD, Brand RA, Leopold SS. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Arch Intern Med* 2010;170:1934-9.
379 Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW, et al. Publication bias in editorial decision making. *JAMA* 2002;287:2825-8.
380 Rochon PA, Sekeres M, Hoey J, Lexchin J, Ferris LE, Moher D, et al. Investigator experiences with financial conflicts of interest in clinical trials. *Trials* 2011;12:9.
381 Steinbrook R. Gag clauses in clinical-trial agreements. *N Engl J Med* 2005;352:2160-2.
382 McCarthy M. Company sought to block paper's publication. *Lancet* 2000;356:1659.
383 Nathan DG, Weatherall DJ. Academic freedom in clinical research. *N Engl J Med* 2002;347:1368-71.
384 Rennie D. Thyroid storm. *JAMA* 1997;277:1238-43.
385 Flanagin A, Fontanarosa PB, DeAngelis CD. Authorship for research groups. *JAMA* 2002;288:3166-8.
386 Ross JS, Hill KP, Egilman DS, Krumholz HM. Guest authorship and ghostwriting in publications related to rofecoxib: a case study of industry documents from rofecoxib litigation. *JAMA* 2008;299:1800-12.
387 Wislar JS, Flanagin A, Fontanarosa PB, DeAngelis CD. Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ* 2011;343:d6128.
388 Gøtzsche PC, Kassirer JP, Woolley KL, Wager E, Jacobs A, Gertel A, et al. What should be done to tackle ghostwriting in the medical literature? *PLoS Med* 2009;6:e1000023.
389 International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. 2010. www.icmje.org/urm_full.pdf.
390 Matheson A. How industry uses the ICMJE guidelines to manipulate authorship--and how they should be revised. *PLoS Med* 2011;8:e1001072.
391 Graf C, Battisti WP, Bridges D, Bruce-Winkler V, Conaty JM, Ellison JM, et al. Good publication practice for communicating company sponsored medical research: the GPP2 guidelines. *BMJ* 2009;339:b4330.
392 Jacobs A, Wager E. European Medical Writers Association (EMWA) guidelines on the role of medical writers in developing peer-reviewed publications. *Curr Med Res Opin* 2005;21:317-21.
393 Wolinsky FD, Vander Weg MW, Howren MB, Jones MP, Martin R, Luger TM, et al. Protocol for a randomized controlled trial to improve cognitive functioning in older adults: the Iowa Healthy and Active Minds Study [protocol]. *BMJ Open* 2011;1:e000218.

394 Chan A-W. Bias, spin, and misreporting: Time for full access to trial protocols and results. *PLoS Med* 2008;5:e230.
395 Lassere M, Johnson K. The power of the protocol. *Lancet* 2002;360:1620-2.
396 Wieseler B, Kerekes MF, Vervoelgyi V, McGauran N, Kaiser T. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. *BMJ* 2012;344:d8141.
397 Gøtzsche PC, Jørgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686.
398 European Medicines Agency. European Medicines Agency policy on access to documents (related to medicinal products for human and veterinary use) (EMA/110196/2006). 2010. www.ema.europa.eu/docs/en_GB/document_library/Other/2010/11/WC500099473.pdf.
399 Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations from the tamiflu experience. *PLoS Med* 2012;9:e1001201.
400 Eichler H-G, Abadie E, Breckenridge A, Leufkens H, Rasi G. Open clinical trial data for all? A view from regulators. *PLoS Med* 2012;9:e1001202.
401 Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council. Sharing publication-related data and materials: responsibilities of authorship in the life sciences. National Academies Press, 2003.
402 Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *Trials* 2010;11:9.
403 Walport M, Brest P. Sharing research data to improve public health. *Lancet* 2011;377:537-9.
404 Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. *Circ Cardiovasc Qual Outcomes* 2012;5:238-40.
405 Vickers AJ. Making raw data more widely available. *BMJ* 2011;342:d2323.
406 The Royal Society Science Policy Centre. Science as an open enterprise. 2012. http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf
407 Summerskill W, Collingridge D, Frankish H. Protocols, probity, and publication. *Lancet* 2009;373:992.
408 Altman D, Furberg C, Grimshaw J, Rothwell P. Trials—using the opportunities of electronic publishing to improve the reporting of randomised trials. *Trials* 2006;7:6.
409 Sharing of materials, methods, and data. 2011. www.plosone.org/static/policies.action.
410 Trials. Instructions for authors. Editorial policies. 2012. www.trialsjournal.com/authors/instructions.
411 National Institutes of Health. Final NIH statement on sharing research data. Feb 26, 2003. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.
412 Laine C, Goodman SN, Griswold ME, Sox HC. Reproducible research: moving toward research the public can really trust. *Ann Intern Med* 2007;146:450-3.
413 BMJ Publishing Group Ltd. Instructions for authors. 2012. http://bmjopen.bmj.com/site/about/guidelines.xhtml.
414 Sugarman J, McCrory DC, Hubal RC. Getting meaningful informed consent from older adults: a structured literature review of empirical research. *J Am Ger Soc* 1998;46:517-24.
415 Paris A, Cracowski JL, Ravanel N, Cornu C, Gueyffier F, Deygas B, et al. [Readability of informed consent forms for subjects participating in biomedical research: updating is required]. *Presse Med* 2005;34:13-8.
416 Southwest Oncology Group. Chemoprevention of prostate cancer with finasteride (Proscar®) Phase III [protocol]. Aug 2001 version. http://swog.org/visitors/pcpt/.
417 Schulz KF, Altman DG, Moher D, the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

and neutral resource that supports and complements efforts of the research enterprise and its key stakeholders.

Universities should insist that their faculties and students are schooled in the ethics of research, their publications feature neither honorific nor ghost authors, their public information offices avoid hype in publicizing findings, and suspect research is promptly and thoroughly investigated. All researchers need to realize that the best scientific practice is produced when, like Darwin, they persistently search for flaws in their arguments. Because inherent variability in biological systems makes it possible for researchers to explore different sets of conditions until the expected (and rewarded) result is obtained, the need for vigilant self-critique may be especially great in research with direct application to human disease. We encourage each branch of science to invest in case studies identifying what went wrong in a selected subset of nonreproducible publications—enlisting social scientists and experts in the respective fields to interview those who were involved (and perhaps examining lab notebooks or redoing statistical analyses), with the hope of deriving general principles for improving science in each field.

"Instances in which scientists detect and address flaws in work constitute evidence of success, not failure."

Industry should publish its failed efforts to reproduce scientific findings and join scientists in the academy in making the case for the importance of scientific work. Scientific associations should continue to communicate science as a way of knowing, and educate their members in ways to more effectively communicate key scientific findings to broader publics. Journals should continue to ask for higher standards of transparency and reproducibility.

We recognize that incentives can backfire. Still, because those such as enhanced social image and forms of public recognition (*10*, *11*) can increase productive social behavior (*12*), we believe that replacing the stigma of retraction with language that lauds reporting of unintended errors in a publication will increase that behavior. Because sustaining a good reputation can incentivize cooperative behavior (*13*), we anticipate that our proposed changes in the review process will not only increase the quality of the final product but also expose efforts to sabotage independent review. To ensure that such incentives not only advance our objectives but above all do no harm, we urge that each be scrutinized and evaluated before being broadly implemented.

Will past be prologue? If science is to enhance its capacities to improve our understanding of ourselves and our world, protect the hard-earned trust and esteem in which society holds it, and preserve its role as a driver of our economy, scientists must safeguard its rigor and reliability in the face of challenges posed by a research ecosystem that is evolving in dramatic and sometimes unsettling ways. To do this, the scientific research community needs to be involved in an ongoing dialogue. We hope that this essay and the report *The Integrity of Science* (*14*), forthcoming in 2015, will serve as catalysts for such a dialogue.

Asked at the close of the U.S. Constitutional Convention of 1787 whether the deliberations had produced a republic or a monarchy, Benjamin Franklin said "A Republic, if you can keep it." Just as preserving a system of government requires ongoing dedication and vigilance, so too does protecting the integrity of science. ■

REFERENCES AND NOTES

1. Trouble at the lab, *The Economist*, 19 October 2013; www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble.
2. R. Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (University of Chicago Press, Chicago, 1973), p. 276.
3. K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge, London, 1963), p. 293.
4. Editorial Board, *Nature* **511**, 5 (2014); www.nature.com/news/stap-retracted-1.15488.
5. B. A. Nosek *et al*., *Science* **348**, 1422 (2015).
6. Institute of Medicine, *Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities* (National Academies Press, Washington, DC, 2014).
7. B. Nosek, J. Spies, M. Motyl, *Perspect. Psychol. Sci.* **7**, 615 (2012).
8. C. Franzoni, G. Scellato, P. Stephan, *Science* **333**, 702 (2011).
9. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *Responsible Science, Volume I: Ensuring the Integrity of the Research Process* (National Academies Press, Washington, DC, 1992).
10. N. Lacetera, M. Macis, *J. Econ. Behav. Organ.* **76**, 225 (2010).
11. D. Karlan, M. McConnell, *J. Econ. Behav. Organ.* **106**, 402 (2014).
12. R. Thaler, C. Sunstein, *Nudge: Improving Decisions About Health, Wealth and Happiness* (Yale Univ. Press, New Haven, CT, 2009).
13. T. Pfeiffer, L. Tran, C. Krumme, D. Rand, *J. R. Soc. Interface* **2012**, rsif20120332 (2012).
14. Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *The Integrity of Science* (National Academies Press, forthcoming). http://www8.nationalacademies.org/cp/projectview.aspx?key=49387.

10.1126/science.aab3847

SCIENTIFIC STANDARDS

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By **B. A. Nosek,* G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni**

POLICY

Transparency, openness, and reproducibility are readily recognized as vital features of science (*1*, *2*). When asked, most scientists embrace these features as disciplinary norms and values (*3*). Therefore, one might expect that these valued features would be routine in daily practice. Yet, a growing body of evidence suggests that this is not the case (*4*–*6*).

A likely culprit for this disconnect is an academic reward system that does not sufficiently incentivize open practices (*7*). In the present reward system, emphasis on innovation may undermine practices that support verification. Too often, publication requirements (whether actual or perceived) fail to encourage transparent, open, and reproducible science (*2*, *4*, *8*, *9*). For example, in a transparent science, both null results and statistically significant results are made available and help others more accurately assess the evidence base for a phenomenon. In the present culture, however, null results are published less frequently than statistically significant results (*10*) and are, therefore, more likely inaccessible and lost in the "file drawer" (*11*).

The situation is a classic collective action problem. Many individual researchers lack

strong incentives to be more transparent, even though the credibility of science would benefit if everyone were more transparent. Unfortunately, there is no centralized means of aligning individual and communal incentives via universal scientific policies and procedures. Universities, granting agencies, and publishers each create different incentives for researchers. With all of this complexity, nudging scientific practices toward greater openness requires complementary and coordinated efforts from all stakeholders.

THE TRANSPARENCY AND OPENNESS PROMOTION GUIDELINES. The Transparency and Openness Promotion (TOP) Committee met at the Center for Open Science in Charlottesville, Virginia, in November 2014 to address one important element of the incentive systems: journals' procedures and policies for publication. The committee consisted of disciplinary leaders, journal editors, funding agency representatives, and disciplinary experts largely from the social and behavioral sciences. By developing shared standards for open practices across journals, we hope to translate scientific norms and values into concrete actions and change the current incentive structures to drive researchers' behavior toward more openness. Although there are some idiosyncratic issues by discipline, we sought to produce guidelines that focus on the commonalities across disciplines.

Standards. There are eight standards in the TOP guidelines; each moves scientific communication toward greater openness. These standards are modular, facilitating adoption in whole or in part. However, they also complement each other, in that commitment to one standard may facilitate adoption of others. Moreover, the guidelines are sensitive to barriers to openness by articulating, for example, a process for exceptions to sharing because of ethical issues, intellectual property concerns, or availability of necessary resources. The complete guidelines are available in the TOP information commons at http://cos.io/top, along with a list of signatories that numbered 86 journals and 26 organizations as of 15 June 2015. The table provides a summary of the guidelines.

First, two standards reward researchers for the time and effort they have spent engaging in open practices. (i) Citation standards extend current article citation norms to data, code, and research materials. Regular and rigorous citation of these materials credit them as original intellectual contributions. (ii) Replication standards recognize the value of replication for independent verification of research results and identify the conditions under which replication studies will be published in the journal. To progress, science needs both innovation and self-correction; replication offers opportunities for self-correction to more efficiently identify promising research directions.

ILLUSTRATION: DAVIDE BONAZZI

Second, four standards describe what openness means across the scientific process so that research can be reproduced and evaluated. Reproducibility increases confidence in results and also allows scholars to learn more about what results do and do not mean. (i) Design standards increase transparency about the research process and reduce vague or incomplete reporting of the methodology. (ii) Research materials standards encourage the provision of all elements of that methodology. (iii) Data sharing standards incentivize authors to make data available in trusted repositories such as Dataverse, Dryad, the Interuniversity Consortium for Political and Social Research (ICPSR), the Open Science Framework, or the Qualitative Data Repository. (iv) Analytic methods standards do the same for the code comprising the statistical models or simulations conducted for the research. Many discipline-specific standards for disclosure exist, particularly for clinical trials and health research more generally (e.g., www.equator-network.org). Many more are emerging for other disciplines, such as those developed by *Psychological Science* (*12*).

Finally, two standards address the values resulting from preregistration. (i) Standards for preregistration of studies facilitate the discovery of research, even unpublished research, by ensuring that the existence of the study is recorded in a public registry. (ii) Preregistration of analysis plans certify the distinction between confirmatory and exploratory research, or what is also called hypothesis-testing versus hypothesis-generating research. Making transparent the distinction between confirmatory and exploratory methods can enhance reproducibility (*3*, *13*, *14*).

Levels. The TOP Committee recognized that not all of the standards are applicable to all journals or all disciplines. Therefore, rather than advocating for a single set of guidelines, the TOP Committee defined

*Corresponding author. E-mail: nosek@virginia.edu
Affiliations for the authors, all of whom are members of the TOP Guidelines Committee, are given in the supplementary materials.

three levels for each standard. Level 1 is designed to have little to no barrier to adoption while also offering an incentive for openness. For example, under the analytic methods (code) sharing standard, authors must state in the text whether and where code is available. Level 2 has stronger expectations for authors but usually avoids adding resource costs to editors or publishers that adopt the standard. In Level 2, journals would require code to be deposited in a trusted repository and check that the link appears in the article and resolves to the correct location. Level 3 is the strongest standard but also may present some barriers to implementation for some journals. For example, the journals *Political Analysis* and *Quarterly Journal of Political Science* require authors to provide their code for review, and editors reproduce the reported analyses publication. In the table, we provide "Level 0" for comparison of common journal policies that do not meet the transparency standards.

Adoption. Defining multiple levels and distinct standards facilitates informed decision-making by journals. It also acknowledges the variation in evolving norms about research transparency. Depending on the discipline or publishing format, some of the standards may not be relevant for a journal. Journal and publisher decisions can be based on many factors—including their readiness to adopt modest to stronger transparency standards for authors, internal journal operations, and disciplinary norms and expectations. For example, in economics, many highly visible journals such as *American Economic Review* have already adopted strong policies requiring data sharing, whereas few psychology journals have comparable requirements.

In this way, the levels are designed to facilitate the gradual adoption of best practices. Journals may begin with a standard that rewards adherence, perhaps as a step toward requiring the practice. For example, *Psychological Science* awards badges for "open data," "open materials," and "preregistration" (*12*), and approximately 25% of accepted articles earned at least one badge in the first year of operation.

The Level 1 guidelines are designed to have minimal effect on journal efficiency and workflow while also having a measurable impact on transparency. Moreover, although higher levels may require greater implementation effort up front, such efforts may benefit publishers and editors and the quality of publications by, for example, re-

Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

ducing time spent on communication with authors and reviewers, improving standards of reporting, increasing detectability of errors before publication, and ensuring that publication-related data are accessible for a long time.

Evaluation and revision. An information commons and support team at the Center for Open Science is available (top@cos.io) to assist journals in selection and adoption of standards and will track adoption across journals. Moreover, adopting journals may suggest revisions that improve the guidelines or make them more flexible or adaptable for the needs of particular subdisciplines.

The present version of the guidelines is not the last word on standards for openness in science. As with any research enterprise, the available empirical evidence will expand with application and use of these guidelines. To reflect this evolutionary process, the guidelines are accompanied by a version number and will be improved as experience with them accumulates.

Conclusion. The journal article is central to the research communication process. Guidelines for authors define what aspects of the research process should be made available to the community to evaluate, critique, reuse, and extend. Scientists recognize the value of transparency, openness, and reproducibility. Improvement of journal policies can help those values become more evident in daily practice and ultimately improve the public trust in science, and science itself. ■

REFERENCES AND NOTES

1. M. McNutt, *Science* **343**, 229 (2014).
2. E. Miguel *et al.*, *Science* **343**, 30 (2014).
3. M. S. Anderson, B. C. Martinson, R. De Vries, *J. Empir. Res. Hum. Res. Ethics* **2**, 3 (2007).
4. J. P. A. Ioannidis, M. R. Munafò, P. Fusar-Poli, B. A. Nosek, S. P. David, *Trends Cogn. Sci.* **18**, 235 (2014).
5. L. K. John, G. Loewenstein, D. Prelec, *Psychol. Sci.* **23**, 524 (2012).
6. E. H. O'Boyle Jr., G. C. Banks, E. Gonzalez-Mule, *J. Manage.* 10.1177/0149206314527133 (2014).
7. B. A. Nosek, J. R. Spies, M. Motyl, *Perspect. Psychol. Sci.* **7**, 615 (2012).
8. J. B. Asendorpf *et al.*, *Eur. J. Pers.* **27**, 108 (2013).
9. J. P. Simmons, L. D. Nelson, U. Simonsohn, *Psychol. Sci.* **22**, 1359 (2011).
10. A. Franco, N. Malhotra, G. Simonovits, *Science* **345**, 1502 (2014).
11. R. Rosenthal, *Psychol. Bull.* **86**, 638 (1979).
12. E. Eich, *Psychol. Sci.* **25**, 3 (2014).
13. E.-J. Wagenmakers, R. Wetzels, D. Borsboom, H. L. van der Maas, R. A. Kievit, *Perspect. Psychol. Sci.* **7**, 632 (2012).
14. C. D. Chambers, *Cortex* **49**, 609 (2013).

ACKNOWLEDGMENTS

This work was supported by the Laura and John Arnold Foundation.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6242/1422/suppl/DC1

10.1126/science.aab2374

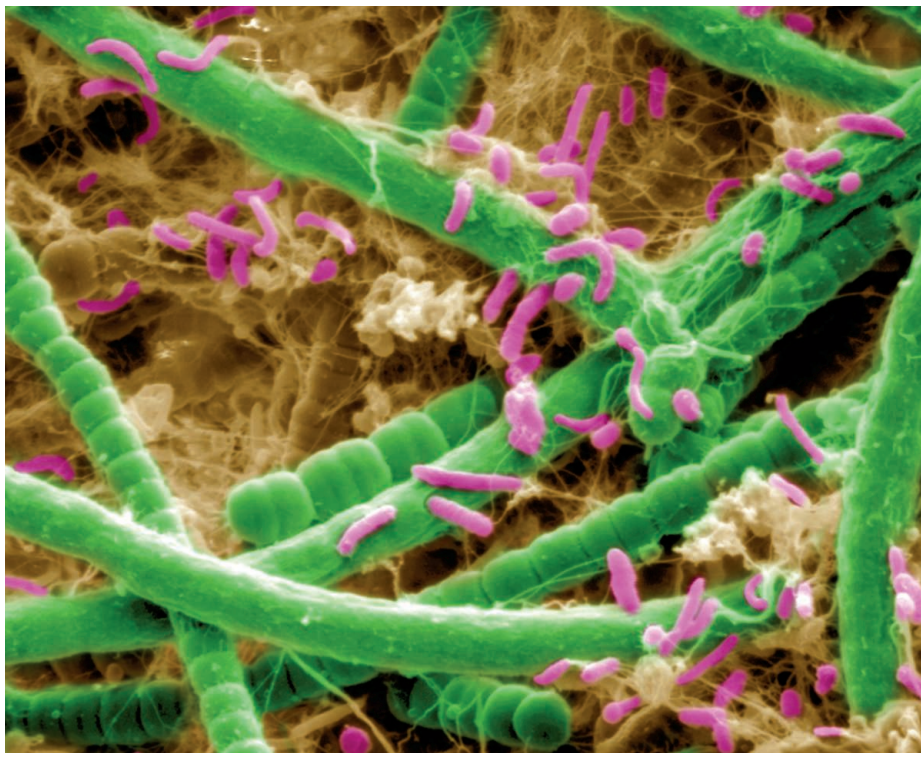


PHOTO: ALICE DOHNALKOVA/PNNL

Learning from nature. Photomicrograph of cyanobacterial-heterotroph microbial consortia derived from a phototrophic microbial mat community from a saline lake. Emerging understanding of cooperative mechanisms in such communities may be helpful in the design of synthetic communities for use in biotechnology.

ECOLOGY

Ecological communities by design

Synthetic ecology requires knowledge of how microbial communities function

By **James K. Fredrickson**

In synthetic ecology, a nascent offshoot of synthetic biology, scientists aim to design and construct microbial communities with desirable properties. Such mixed populations of microorganisms can simultaneously perform otherwise incompatible functions (*1*). Compared with individual organisms, they can also better resist losses in function as a result of environmental perturbation or invasion by other species (*2*). Synthetic ecology may thus be a promising approach for developing robust, stable biotechnological processes, such as the conversion of cellulosic biomass to biofuels (*3*). However, achieving this will require detailed knowledge of the principles that guide the structure and function of microbial communities (see the image).

Recent work with synthetic communities is shedding light on microbial interactions that may lead to new principles for community design and engineering. In game theory, cooperators provide publicly available goods that benefit all, whereas cheaters exploit those goods without reciprocation. The tragedy of the commons predicts that cheaters are more fit than cooperators, eventually destroying the cooperation. Yet, this is not borne out by observations. For example, using a synthetic consortium of genetically modified yeast to represent cooperators and cheaters, Waite and Shou (*4*) found that, although initially less fit than cheaters, cooperators rapidly dominated in a fraction of the cultures. The evolved cooperators harbored mutations allowing them to grow at much lower nutrient concentrations than their ancestor. This suggests that the tragedy of the commons can be avoided

Science

Promoting an open research culture

B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson and T. Yarkoni

Science **348** (6242), 1422-1425.
DOI: 10.1126/science.aab2374

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

shortcut, investing in magic bullets such as vaccines, antiretrovirals, and bednets that could be distributed without a well functioning health-care system, and even in war zones.[8] These programmes saved millions of lives, but failed to launch the anticipated virtuous cycle of development, poverty reduction, and improved health systems. Many African countries remain deeply impoverished, and although child mortality has plummeted, weak health systems remain ill-equipped to address soaring numbers of deaths and disability from heart disease, cancer, and other non-communicable diseases.[9]

Bollyky and colleagues' findings suggest a possible reason for this dispiriting reality: many of the countries with the worst health systems are governed by corrupt autocrats who retain power by force and can ignore the welfare of their people without repercussions.

To date, the global development community has skirted the complications of politics, instead emphasising medical programmes and the empowerment of women, the LGBT+ (lesbian, gay, bisexual, transgender, and all other identities) community, and people with disabilities—ie, people who do not directly challenge government power. This emphasis is sometimes justified by the longstanding assumption that in poor countries, especially, dictators are better at getting things done because they can ignore the demands of petty, competing constituencies.[10] However, Bollyky's findings support the theory that dictators might themselves be a cause of poverty and illness, and that democrats, however befuddled and disorganised, better serve their people.

Unfortunately, the rights of pro-democracy activists are routinely violated with impunity, especially in those African countries most beset by poverty and ill health.[11] Recent elections in Kenya, Uganda, Ethiopia, Rwanda, the Democratic Republic of the Congo, Gabon, Cameroon, and Zimbabwe were all marred by credible rigging allegations, which donors, in most cases, dismissed. When two opposition members of parliament in donor darling Uganda were beaten and crippled by security forces inside parliament in September, 2017, not one international donor or human rights organisation spoke out.[12]

Global health advocacy groups need to do more than clamour for more funding and occasionally bemoan corruption. They need to call on Washington (USA), Brussels (Belgium), London (UK), and other donors to impose sanctions on dictators, including those who cooperate with western military aims. As Rudolf Vichow, one of the pioneers of modern public health, wrote after witnessing the ravages of typhus on the oppressed peasants of Silesia, "politics is nothing but medicine at a larger scale".[13]

Helen Epstein
Bard College, Annandale-on-Hudson, NY 12504, USA
hepstein@bard.edu

I declare no competing interests.

1 Bollyky TJ, Templin T, Cohen M, et al. The relationships between democratic experience, adult health, and cause-specific mortality in 170 countries between 1980 and 2016: an observational analysis. *Lancet* 2019; published online March 13. http://dx.doi.org/10.1016/S0140-6736(19)30235-1.
2 Qian N, Yanagizawa-Drott D. The strategic determinants of US human rights reporting: evidence from the Cold War. *JEEA* 2009; **7:** 446–57.
3 Smith A. An inquiry into the nature and causes of the wealth of nations. London: W Strahan & T Cadell, 1776.
4 Szreter S. The population health approach in historical perspective. *Am J Public Health* 2003; **93:** 421–31.
5 Almond D, Chay KY, Greenston M. Civil rights, the war on poverty, and black-white convergence in infant mortality in the rural south and Mississippi. MIT Department of Economics Working Paper No. 07-04. Dec 31, 2006. https://ssrn.com/abstract=961021 (accessed Feb 12, 2019).
6 Bollyky TJ. Plagues and the paradox of progress: why the world is getting healthier in worrisome ways. Cambridge, MA: MIT Press, 2018.
7 Deaton A. The great escape: health, wealth and the origins of inequality. Princeton, NJ: Princeton University Press, 2015.
8 Fifield A. A mighty purpose. How Jim Grant sold the world on saving its children. New York: Other Press, 2015.
9 Nyaaba GN, Stronks K, de-Graft Aikins A, Kengne AP, Agyemang C. Tracing Africa's progress towards implementing the non-communicable diseases global action plan 2013–2020: a synthesis of WHO country profile reports. *BMC Public Health* 2017; **17:** 297.
10 Przeworski A, Alvarez ME, Cheibub JA, Limongi F. Democracy and development: political institutions and well-being in the world, 1950–1990. Cambridge: Cambridge University Press, 2000.
11 Epstein H. Another fine mess: America, Uganda and the war on terror. New York: Columbia Global Reports, 2017.
12 Epstein H. The US turns a blind eye to Uganda's assault on democracy. *The Nation* July 20, 2018.
13 Virchow R. Der Armenarzt. *Medicinische Reform* 1848; **18:** 125–27.

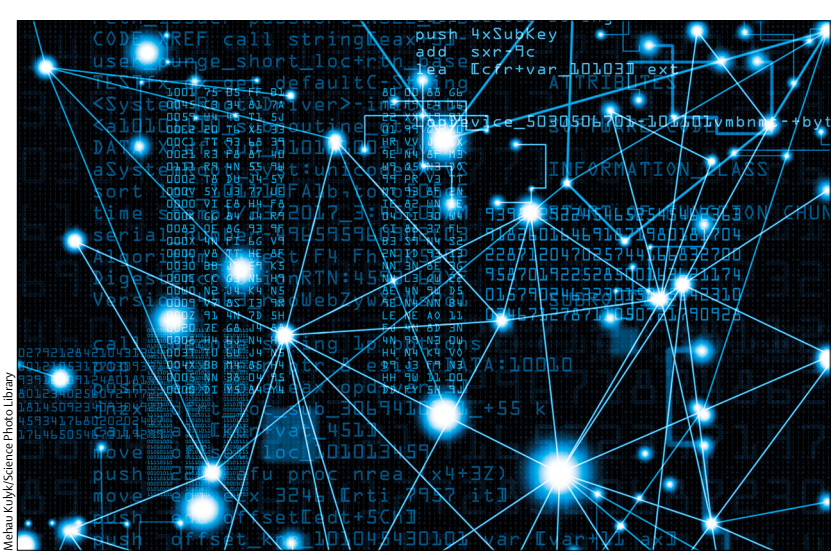
Reporting of artificial intelligence prediction models

Data-driven technologies that form the basis of the digital health-care revolution provide potentially important opportunities to deliver improvements in individual care and to advance innovation in medical research. Digital health technologies include mobile devices and health apps (m-health), e-health

technology, and intelligent monitoring. Behind the digital health revolution are also methodological advancements using artificial intelligence and machine learning techniques. Artificial intelligence, which encompasses machine learning, is the scientific discipline that uses computer algorithms to learn from data, to help identify patterns in data, and make predictions. A key feature underpinning the excitement behind artificial intelligence and machine learning is their potential to analyse large and complex data structures to create prediction models that personalise and improve diagnosis, prognosis, monitoring, and administration of treatments, with the aim of improving individual health outcomes. Prediction models to support clinical decision making have existed for decades, and these include well known tools such as the Framingham Risk Score,[1] QRISK3,[2] Model for End-stage Liver Disease,[3] ABCD² score,[4] and the Nottingham Prognostic Index.[5] Health-care professionals, medical researchers, policy makers, guideline developers, patients, and members of the general public are all potential users of prediction models. The number of prediction model studies is increasing rapidly, with hundreds of different models being developed for some of the same targeted populations and outcomes.[6,7]

In response to this rapid growth, as well as concern about incomplete reporting of prediction model studies,[8,9] the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement was published in 2015. TRIPOD provides guidance on the key items to report when describing studies developing, evaluating (or validating), or updating clinical prediction models.[10,11] Although TRIPOD aims primarily to improve reporting, it also leads to more comprehensive understanding, conduct, and analysis of prediction model studies, ensuring that prediction models can be picked up by subsequent researchers and users to be studied further and used to guide health care, thus encouraging reproducible research and reduce research waste.

For more on the **TRIPOD statement** see https://www.tripod-statement.org



Mehau Kulyk/Science Photo Library

Application of artificial intelligence and machine learning is often championed as a solution to prediction problems that involve big datasets, large numbers of predictors, and different sources and types of data (such as free-text notes). However, concerns have been raised that artificial intelligence in clinical medicine is overhyped and, if not used with proper guidance, knowledge, or expertise, has methodological shortcomings, poor transparency, and poor reproducibility.[12] Methodological concerns include an often incorrect focus on classification over prediction, overfitting (whereby too many predictors or features are included for the sample size), lack of robust assessment of predictive accuracy when used with other data than those from which they were developed (validation), weak and unbiased comparison with simpler modelling approaches, and lack of transparency of the artificial intelligence and machine learning algorithm, which limits independent evaluation.

There is a strong need that machine learning techniques adhere to established methodological standards already defined in prediction model research. Fair and neutral evaluations and comparisons against these existing prediction model approaches must be done. In particular, regression-based prediction models, for which an equation links a set of predictors to an outcome, can easily be recorded so that they can be implemented by others. In contrast, prediction models based on machine learning are much more reliant on computers for implementation of the underlying model, which is often labelled as a black box. This inherent complexity behind machine learning obfuscates interpretation or face-value assessment of the prediction model algorithm and hampers, or at least requires thought on, how best to make the machine learning algorithm available to others for independent validation or to implement it into the clinical workflow.

Clearly, the consequences of making a wrong or inaccurate prediction are substantial for the clinical

application of a machine learning prediction model, such as the deep learning models for detection of stroke or wrist fractures approved by the US Food and Drug Administration.[13] Therefore, the clinical community must not get mesmerised by the artificial intelligence and machine learning revolution, and artificial intelligence and machine learning prediction models must be appropriately developed, evaluated, and—if needed—tailored to different situations before they are used in daily medical practice. Complete and transparent reporting of the key aspects of machine learning prediction model studies in the light of existing evidence is vital.

Although many aspects of the TRIPOD statement are applicable to prediction model studies using artificial intelligence and machine learning methods,[10] its uptake by the artificial intelligence and machine learning communities has not been high. Possible reasons for this include subtle differences in terminology, or a perceived lack of relevance because TRIPOD focuses on regression-based prediction model approaches. Also, TRIPOD's explanation and elaboration document, which provides examples of good reporting to help authors, focuses on models developed using regression.[11]

We therefore announce a new initiative to develop a version of the TRIPOD statement specific to machine learning (TRIPOD-ML). The aim of TRIPOD-ML will be to focus on the introduction of machine learning prediction algorithms, building on a long and established methodology of prediction research, while harmonising terminology. To make this new guideline as usable as possible, we would like to invite interested individuals from the machine learning community, particularly those with an interest in machine learning and artificial intelligence applications in health care, to contribute.

*Gary S Collins, Karel G M Moons
Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK (GSC); and Julius Center for Health Sciences and Primary Care, and Cochrane Netherlands, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands (KGMM)
gary.collins@csm.ox.ac.uk

We declare no competing interests.

1 Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; **97:** 1837–47.
2 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; **357:** j2099.
3 Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 2000; **31:** 864–71.
4 Johnston SC, Rothwell PM, Nguyen-Huynh MN, et al. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet* 2007; **369:** 283–92.
5 van Gorp MJ, Steyerberg EW, van der Graaf Y. Decision guidelines for prophylactic replacement of Bjork-Shiley convexo-concave heart valves: impact on clinical practice. *Circulation* 2004; **109:** 2092–96.
6 Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016; **353:** i2416.
7 Tangri N, Kitsios GD, Inker LA, et al. Risk prediction models for patients with chronic kidney disease: a systematic review. *Ann Intern Med* 2013; **158:** 596–603.
8 Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012; **9:** e1001221.
9 Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Meth* 2014; **14:** 40.
10 Collins GS, Reitsma JB, Altman DG., Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis: the TRIPOD statement. *Ann Intern Med* 2015; **162:** 55–63.
11 Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162:** W1–73.
12 Chen JH, Asch SM. Machine Learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med* 2017; **376:** 2507–09.
13 Ratner M. FDA backs clinician-free AI imaging diagnostic tools. *Nat Biotechnol* 2018; **36:** 673–74.

Opioid analgesics pass the acid test

Understanding of the neurobiology of pain continues to increase at a remarkable pace, raising optimism for improvements in pain therapy. Physicians can choose from multiple classes of medications to treat acute or chronic pain that can often be highly effective. Many patients, however, do not have their pain well controlled with available drugs, and these drugs often have severe side-effects that can ultimately diminish quality of life and potentially result in serious harm.[1] Hence, the need for new medicines for treatment of pain remains very high.[2]

Opioids are effective in the treatment of moderate-to-severe acute pain conditions including postoperative pain, inflammatory pain, trauma-related pain, and cancer pain. Although some patients with chronic non-malignant pain respond well to opioids, the medical community is increasingly aware of the risks that come with opioid use for chronic pain.[1]