Evidence Generation for Genomic Diagnostic Test Development

Session: Evidence

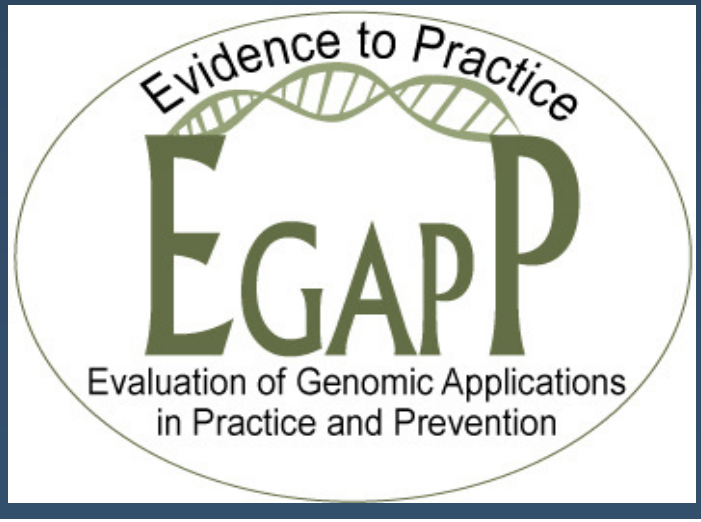
Ned Calonge, MD, MPH
Chair, EGAPP Working Group

Objectives

- Provide brief review of EGAPP Working Group methods
- Discuss EGAPP recommendations for
  - Lynch syndrome
  - Breast cancer gene expression profiling

EGAPP

Evaluation of
Genomic
Applications in
Practice and
Prevention



- CDC-funded initiative
- Non-regulatory
- Independent, non-federal, multidisciplinary Working Group
- Integrate existing processes for evaluation and appraisal
- Minimize conflicts of interest
- Evidence-based, transparent, and publicly accountable

EGAPP methods

- Integrate knowledge and experience from existing processes
  - Systematic review process from ACCE
  - Assessment of quality of individual studies, adequacy of overall evidence, level of certainty, and magnitude of net benefit from USPSTF and EPCs
  - Contextual issues from GRADE
- Provide recommendations with clear linkage to the evidence
- Identify gaps - inform research agenda

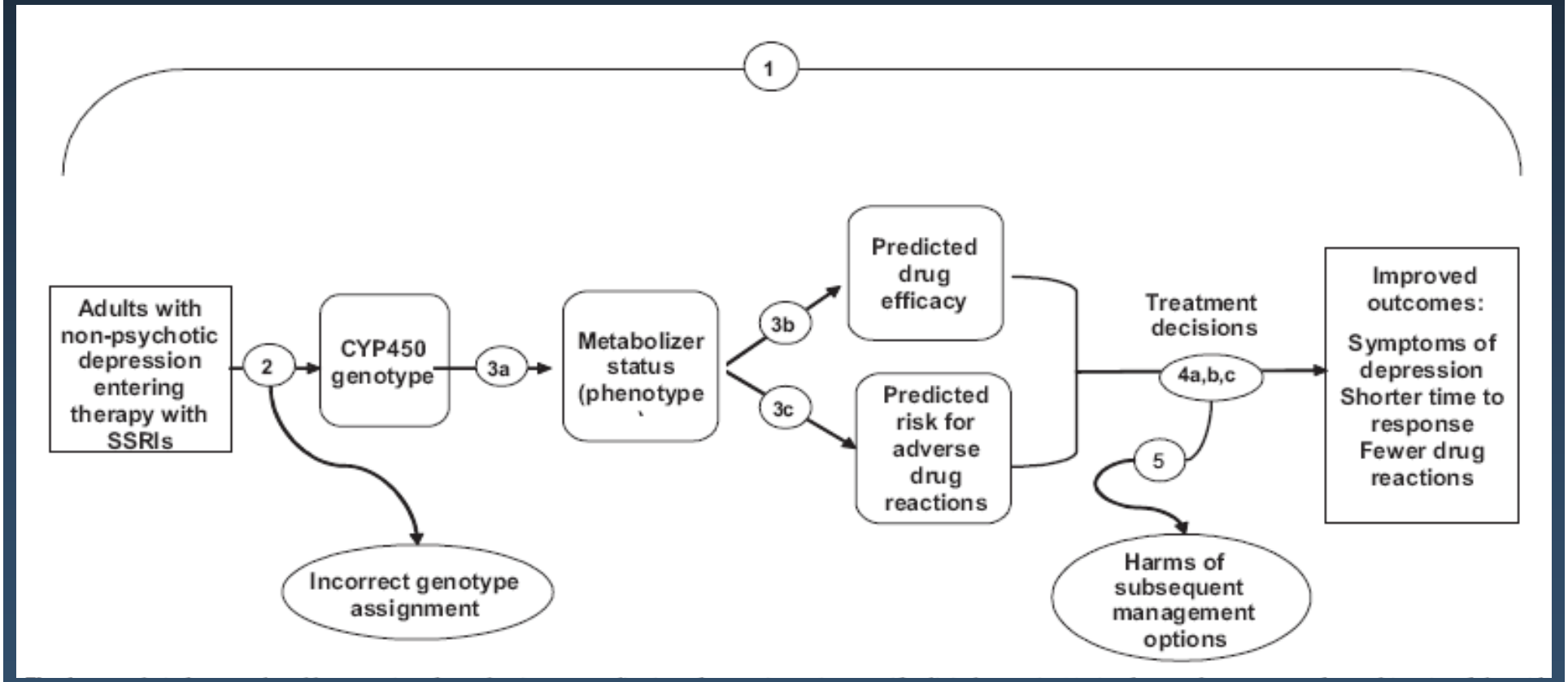
Steps in the EWG process

- Select topic/genomic application for evaluation
- Define clinical scenario (diagnosis, disease screening, risk assessment, prognosis, pharmacogenetics)
- Create analytic framework of key questions to guide the evidence review
- Find, synthesize and evaluate quality and adequacy of existing literature
- Determine net benefit (benefit minus harms) of test application
- Create a recommendation based on the certainty of net benefit

Analytic Framework and evidence chain

- If there is no overarching (direct) evidence (Key Question 1), the framework is used to develop a chain of evidence that can lead to a recommendation
- Questions address (from ACCE):
  - Analytic validity
  - Clinical validity
  - Clinical utility

Analytic framework



Key Questions in evidence chain

KQ2: Analytic validity (technical performance)

- Analytic sensitivity/specificity, reliability, assay robustness

KQ3: Clinical validity

- Ability to identify or predict the disorder of interest: clinical sensitivity, specificity, predictive value relating to expression/phenotype

KQ4&5: Clinical utility

- Balance of benefits and harms with clinical use: efficacy/effectiveness and net benefit

Grade the quality of the evidence for the key questions

- Quality accounts for the hierarchial level of the study design as well as study flaws or threats to internal validity
- Evidence is classified as:
  - Convincing (observed effect is likely to be real)
  - Adequate (a higher risk that the effect may be influenced by study flaws)
  - Inadequate (too many flaws to confidently assign the results to the factors under study)

Determination of net benefit

- Involves balancing benefits or potential benefits with harms or potential harms
- Often involves comparing harms and benefits that are very disparate in terms of health or value
- Classified as small, moderate or substantial

Determination of certainty of net benefit

- Certainty is based on the overall assessment of the evidence
- Certainty relates to the risks of being wrong
- High certainty—unlikely that future research will change the recommendation
- Moderate certainty—based on some questions about the evidence, an increased risk that future research will change the recommendation, but still acceptable
- Low certainty—evidence is inadequate to make a recommendation

Translation into a recommendation

- High or moderate certainty of a small to substantial net benefit: Recommend for…
- High or moderate certainty of a zero benefit or net harm: Recommend against…
- Low certainty: insufficient evidence

Classifying insufficient evidence conclusions

- Based on contextual factors and/or decision modeling topics with a conclusion of insufficient evidence may be classified as:
  - neutral—not possible to predict what future research will find
  - discouraging—discourage use until specific gaps in knowledge are filled or not likely to ever meet evidentiary standards
  - encouraging—likely to meet evidentiary standards with further study or reasonable to use in limited situations while awaiting further evidence

Genetic testing for Lynch syndrome

- The EWP “found sufficient evidence to recommend offering genetic testing for Lynch syndrome to individuals with newly diagnosed colorectal cancer to reduce mortality and morbidity in relatives”

Analytic validity

- Microsatellite instability testing (MSI)
  - Analytic performance high by CAP external proficiency testing with some deficiencies
- Immunohistochemical testing (IHC)
  - Lynch mismatch-repair (MMR) gene proteins not currently subject to CAP testing, but IHC for other proteins is; assume adequate AV
- BRAF mutation testing
  - Analytic performance should be high (single mutation; also, has high clinical validity)

Clinical validity

- MSI (adequate evidence, 11 studies/150 patients):
  - Sensitivity 77-89%, specificity 90%
- IHC (adequate evidence, 9 studies/149 patients)
  - Sensitivity 83%, specificity 90%
- BRAF (adequate evidence, 3 studies/43 patients)
  - Sensitivity 69%, specificity 100%)

Clinical utility– probands

- Insufficient evidence to support differential treatment options based on Lynch syndrome
- Small body of evidence suggests MSI-high tumors relatively resistant to 5FU and more sensitive to irinotecan

Clinical utility- 1st and 2nd degree relatives

- 53-100% uptake in colonoscopy in test positive relatives (7 studies), low harms
- 62% reduction in CRC and significant reduction in CRC mortality in relatives with Lynch syndrome mutations who chose colonic surveillance (n=252)
- 73% mortality reduction in 2,788 persons (147 Lynch syndrome families)
- Similar indirect evidence for women and ovarian/endometrial cancer screening uptake and decreased incidence with surgery

Breast cancer tumor gene expression profiles

- Review included Oncotype DX, MammaPrint and H:I ratio test
- The EWG “found insufficient evidence to make a recommendation for or against the use of tumor gene expression profiles to improve outcomes in defined populations of women with breast cancer”

Breast cancer tumor gene expression profiles

- The EWG “found preliminary evidence of potential benefit of testing results to some women who face decisions about treatment options (reduced adverse events due to low risk women avoiding chemotherapy) but could not rule out the potential for harm for others (breast cancer recurrence that could have been prevented).”

Analytic validity

- Some data on technical performance were available but estimates of analytic sensitivity and specificity could not be made
- Testing initially failed on 14.5-19% of fresh samples
- Evidence was judged as inadequate

Clinical validity

- Oncotype DX
  - Adequate evidence (3 studies, 1 higher quality/ 1,100 patients) supports association with 10-year distant metastasis risk (RR=2.5 high vs. low risk)
- Mammaprint
  - Adequate evidence (2 studies/369 patients) supports association with metastasis at 5-year (RR=7.8) and 10-year (RR=3.3) but unclear relative efficacy in ER+ (better) and ER- (N.S) patients

Clinical utility

- MammaPrint
  - No studies of clinical utility

Clinical utility

- Oncotype DX
  - Retrospective analysis of one arm of prospective clinical trial: high risk women had 27% reduction in 10-year recurrence rate with chemo (less than optimal study design)
  - Many would have been offered chemo without testing
  - 75% of results are intermediate or low risk, where estimates of recurrence risk have wide and overlapping confidence intervals

Clinical utilitiy

- Potential for significant harms (recurrence and/or death)for small number of low and intermediate risk women who might benefit from chemo but forego it
- More women likely to benefit (avoid unnecessary chemo)
- No data on use in women with high risk on conventional assessment but low on Oncotype DX

Clinical utility

- Encouraging indirect evidence for Oncotype DX, plausible potential use for MammaPrint
- TAILORx and MINDACT trials underway

Common research gaps

- Analytic validity
  - often missing information due to proprietary issues or lab-developed tests
- Clinical validity
  - often based on testing subjects with potential sources of bias
- Clinical utility
  - Major source of insufficient conclusion
  - Few RCTs
  - Observational studies have trouble with bias
  - Recommendations based on observational studies run higher risk of being wrong

Other completed EGAPP recommendations

- Insufficient evidence to recommend for or against CYP450 testing to inform SSRI therapy, discourage use until further clinical trials are completed
- Insufficient evidence to recommend for or against UG1A1 genotyping in CRC patients to be treated with irinotecan with the intent of lowering the dose to avoid severe drug reactions
- Accepted for publication:
  - Cardiogenomic profiling
  - Factor V Leiden/prothrombin G20210A

For more information

- EGAPP Working Group website

  http://www.egappreviews.org