New Test Evaluation – the Principles

Graeme P Young.

EWG chairs:
Bob Bresalier, Carlo Senore, Graeme Young
Scope of 2016 recommendations

To develop practical advice on how best to compare “new” with proven screening tests, the ideal context, the informative endpoints and the appropriate study design.

Recommendations for a Step-Wise Comparative Approach to the Evaluation of New Screening Tests for Colorectal Cancer

Graeme P. Young, MD, FRACP, FTSE, AGAF; Carlo Senore, MD, MSc; Jack S. Mandel, PhD, MPH; James E. Allison, MD, FACP, AGAF; Wendy S. Atkin, MPH, PhD; Robert Benamouzig, MD, PhD; Patrick M. M. Bossuyt, PhD; Mahinda De Silva, MB, BS, FRACP; Lydia Guittet, MD, PhD; Stephen P. Halloran, MBE, FRCPa; Ulrike Haug, PhD; Geir Hoff, MB, ChB, PhD; Steven H. Itzkowitz, MD, FACP, FACC, AGAF; Marcis Leja, MD, MBA, PhD, AGAF; Bernard Levin, MB, BCh, FACP; Gerrit A. Meijer, MD, PhD; Colm A. O’Morain, MD; Susan Parry, MBChB, FRACP; Linda Rabeneck, MD, MPH, FRCPC; Paul Rozen, MD; Hiroshi Saito, MD, PhD; Robert E. Schoen, MD, MPH; Helen E. Seaman, BSc, PhD; Robert J. C. Steele, MD, FRCS; Joseph J. Y. Sung, MD, PhD; and Sidney J. Winawer, MD

Cancer 2016; 122(6):826-39
What has transpired?

• New developments in biomarker technologies.
• Widespread implementation of organised population screening that makes test evaluation difficult in intended-to-use populations.
• Differing goals of screening programs around the world.
• The evidence required by regulatory authorities differs from that of health-care providers.
• Omissions and updating:
  • Algorithm complexity and associated challenges were not included.
  • The biomarker section was very basic and did not allow for marker panels.
  • Some other elements as covered in the April meeting.
Guiding Principles – 2016 and planning

• Eight guiding principles which arose from a consensus approach and a literature review.

• In planning the update, it was recognised that the Guiding Principles warranted critical review.
  • Undertake a review based on the Glaser and Delphi approaches to achieving consensus.
  • Approach was as previously undertaken
  • Consensus of 80% agreement was desired (5 grade scale)
Guiding principles, n=10

1. Screening aims to reduce mortality through early detection and treatment of CRC and/or reduce incidence through detection of important preinvasive lesions, without adversely affecting the health status of those who participate and in a feasible and cost-effective manner.

2. The screening test is just one event in a complex multistep process that includes initial and ongoing engagement of the agreed targeted subgroup of the population, testing, validation, communication and treatment.

3. Population randomized controlled trials (RCTs) with mortality as primary outcome set the expectations for performance of new tests. A new test is well understood if compared to an RCT-proven test (the comparator) and effectiveness trials point to intermediate/surrogate endpoints that predict potential for reducing mortality since they can be directly related to the main outcome (mortality and or incidence) of interest.

4. New tests can be assessed in parallel to an existing RCT-proven test at all steps in the screening process from population invitation to key outcomes/measures. The impact of the new test can be predicted from a direct comparison with a proven test using intermediate/surrogate endpoints such as test accuracy, true-positive rate and a shift to earlier stage, as long as the association of these outcomes to the expected health impact has been demonstrated.

5. New screening tests might detect a neoplasm-dependent biology that is different from proven tests. As a consequence, the value of treatment and benefit to survival might not be the same. Concordance between new and established tests is important, together with detection of cases with the new test which would not have been detected with the old test. However, the fact that structural detection of adenomas by endoscopic means leads to reduction in CRC incidence and mortality, indicates that it is unlikely that a shift in treatability will prove to be a real issue of concern.

6. In two-step screening, the screening test selects participants who proceed to diagnostic verification by colonoscopy, as a positive test increases the likelihood of neoplasia being present. No simple test for use in two-step screening will be as accurate as colonoscopy.

7. Reasoned consideration of the evaluation process of a new test gives rise to the concept of a phased or sequential evaluation starting with relative accuracy in moderate-scale studies followed by screening pathway evaluation requiring larger intended-use populations. [repositioned & separated]

8. It is not ethically justifiable to proceed to study acceptability of a test to invitees or to other screening program outcomes without studies that indicate that the new test is at least as accurate as a proven comparator test. The minimum standard now required is that of FIT.

9. A simple screening test should have a flexible endpoint that enables a screening program to select the accuracy that achieves desired program outcomes, with feasible workload. [addition]

10. New tests must be clearly defined with provision of adequate technical details, quality assurance procedures and performance standards and will be required to meet regulatory requirements of an existing jurisdiction. A test should be performed according to professional standards or manufacturer’s instructions, following appropriate training and in accordance with defined quality control and assessment procedures and with good laboratory and clinical practice. REWORD

Tested using a modified Delphi approach in 47 experts from WEO.

So far, 2 rounds have been undertaken.
Results – rounds 1 & 2

Proportion agree (score 4 or 5)

Proportion

Statement number

1  2  3  4  5  6  7  8  9  10

Proportion agree (score 4 or 5)

Proportion

Statement number

1  2  3  4  5  6  7  8  9  10  11  12
For a test to succeed – a new principle

11. To succeed in programs and guidelines, application at key steps in the multistep pathway in an intended-use population is required. 89%

- Simply having identified accuracy would not be enough
- **Endpoints:** engagement in screening, test failure rates, workloads, and determination of cost effectiveness as well as surrogate measures for mortality benefit (shift to earlier stage) and incidence.
- Comparative effectiveness RCTs are ideal for such purposes.
- Longitudinal as well as cross-sectional data are ideal.
For a test to succeed — a new principle

11. To succeed in programs and guidelines, application at key steps in the multistep pathway in an intended-use population is required. 89%

- Simply having identified accuracy would not be enough
- **Endpoints**: engagement in screening, test failure rates, workloads, and determination of cost effectiveness as well as surrogate measures for mortality benefit (shift to earlier stage) and incidence.
- Comparative effectiveness RCTs are ideal for such purposes.
- Longitudinal as well as cross-sectional data are ideal.
- Dialogue continues on what endpoints might be adequate. An emerging controversy: Is earlier stage at diagnosis enough to predict mortality benefit?
- However, this principle was considered to be a high hurdle for regulatory approval.
Evaluation is necessarily phased

7. The process of evaluation should follow a phased (sequential) evaluation starting with moderate-scale studies.
   • Early studies might be cohort or case-control studies including subjects with and without neoplasia. Initial estimates of clinical performance characteristics and analytical specifications will be obtained. (Phases 1 and 2).
   • If results suggest that the test compares well with the desired standard, evaluation should proceed to screening pathway evaluation requiring larger intended-use populations. The latter should be prospective.

• Concept was fine but more detail was requested.
• For resolution: When does test threshold for positivity get set?
• Endpoint?: Is not mortality benefit beyond the scope of the test?
Standard comes from RCTs and comparator test

3. Population RCTs with mortality and/or incidence reduction as the outcome set the expectations for performance. Effectiveness of a new test may be predicted when compared to a reference standard (comparator or index test) in a paired design where the comparator’s effectiveness is supported by evidence.  

- Revisions made to avoid the implication that gFOBT were the standard and the comparator.
- Debate as to whether FIT should be specified but there are many possible configurations of FIT.
  - Discussion ongoing re mtsDNA and colonoscopy
Test defines likelihood for neoplasia

6. In two-step screening based on first performing a non-invasive test, the test should identify those participants with an increased likelihood of neoplasia, who would then be expected to proceed to diagnostic verification (colonoscopy).

- Restructured and simplified for clarity.
- Original version was a principle for the screening pathway which turned the focus from test evaluation to diagnostic approach.
- Controversial sentence indicating that a noninvasive test would not be expected to be as good as colonoscopy was removed.
Worthwhile target lesions - new

12. For screening to be effective, it requires a noninvasive test in which the analytical target molecule is used to identify an increased likelihood of relevant colorectal neoplasms.

   • The goal of screening is to: **a)** find curable cancer and **b)** remove relevant preinvasive lesions (PIL).
   • Not all adenomas and serrated lesions need to be detected. Currently, we are most interested in those classed as “advanced”.
   • *However, the natural history of PIL is not well understood and more research is needed to clarify how best to characterise those that are the most important to detect.*

   • Biomarkers might provide the best characterization of PIL.
   • It is important to avoid overdiagnosis because of the load on surveillance.
Compare using intermediate endpoints

4. Performance ...can be assessed in parallel to an existing test of proven efficacy at any step in the screening process .... Intermediate endpoints known to reliably and consistently predict potential for reducing mortality and/or incidence, should be used. Such include estimates of test accuracy for cancer and precancer lesions (especially sensitivity and specificity) but also shift to earlier stage.

- This statement has been slightly refocussed to point clearly to comparison with a reference standard at any step and how intermediate measures can be used to guide the researcher as to its potential value.

- **An emerging controversy: Is earlier stage at diagnosis enough to predict mortality benefit?**
Differing biology and survival benefit

5. Non-invasive tests targeting new biomarkers, might detect lesions with different biology and natural history from those detected by established tests. Survival benefit might differ.  
   - Cancers might respond differently to treatment, regardless of stage.
   - PIL might have a different risk profile (genotype or phenotype).
   - Exploring concordance between the new test and the reference standard test will clarify which lesions are newly-detected and which are missed.
   - The value of treating lesions solely detected by new technologies should be determined; are they are unresponsive, of low risk of progression or indolent?
   - Clarity of the original wording was an issue.
Adjustable thresholds are essential

9. A screening test with an adjustable threshold (or algorithm) for test positivity enables a program to choose the test accuracy parameters that best achieve desired program outcomes.

- Regulatory approval processes should recognise the capacity for adjustment of the positivity-threshold when different performance characteristics are necessary to achieve screening program goals.

- “Adjustable” was a much better word than flexible.

- This view can be at variance with regulatory bodies.
Ethics of proceeding – standards?

8. It is not ethically justifiable to proceed to study a test in a typical intended-use population without studies that indicate that the new test is at least as accurate ... as a proven comparator test.

- The minimum standard ... is that of FIT configured to return as good, or better, accuracy than that attained in the population RCTs, with mortality reduction as the primary outcome.
- Exceptions: as an adjunct to FIT, if screening by any other means is refused.
- **Crucial issue:** What should the FIT standard be?
- **US CMS guidelines** about criteria for blood-based CRC screening tests recommend at least 74% sensitivity for CRC (based on FIT) and 90% specificity (no neoplasia, based on the Cologuard test).
Test analytical issues

10. Since the analytical performance characteristics of a test have ramifications for interpretation of results and clinical outcomes, a prerequisite to assessment of the clinical value is the documentation of these characteristics.

• Expert feedback required better wording. Key points in new Revision:
  • Documentation of the analytical performance characteristics to be provided by manufacturers of tests in their instructions for use.
  • Evaluations of the characteristics should be performed using the techniques documented by the Clinical and Laboratory Standards Institute.
  • Researchers should follow such protocols and ensure that standards are attained or surpassed.
  • ISO 15189 specifies the requirements for competence and quality, defines all aspects of the performance of a test. Sites performing tests should be accredited to ISO 15189 standards.
  • Tests should also meet relevant regulatory requirements appropriate to the region of its use.
Conclusions

• Consensus now achieved on the principles
• They remain open for comment and fine-tuning of wording is yet to be completed with the panel.
• Accompanying explanatory text will be provided.
• Eventual publication will include additional materials presented as part of the webinars.

• Institutional and/or personal relationships with:
  • Clinical Genomics Inc.
  • Eiken Chemical Company
ENDO 2022
Connecting the World of Endoscopy
3rd World Congress of GI Endoscopy

In cooperation with
103rd Congress of the
Japan Gastroenterological Endoscopy Society

May 13-15, 2022
Kyoto, Japan
Kyoto International Conference Center & Grand Prince Hotel

www.worldendo.org

SAVE THE DATE!