



WEO

The voice of world
endoscopy

Evaluation of New Tests Roundtable

Robert S. Bresalier M.D.

Professor of Medicine

Distinguished Professor in Gastrointestinal Oncology

MD Anderson Cancer Center



In this report, we revise the previous WEO principles that are fundamental to evaluation of new *non-invasive* tests. This accounts for emerging knowledge, challenges faced as screening becomes much more widespread [WEO2018] and opportunities. Our goal of providing an efficient and feasible, but rigorous strategy to evaluate emerging alternative (hereafter “new”) non-invasive tests and to compare them with the proven/established non-invasive screening tests, remains.



Intended Use: One Size May Not Fit All

Clinical Application

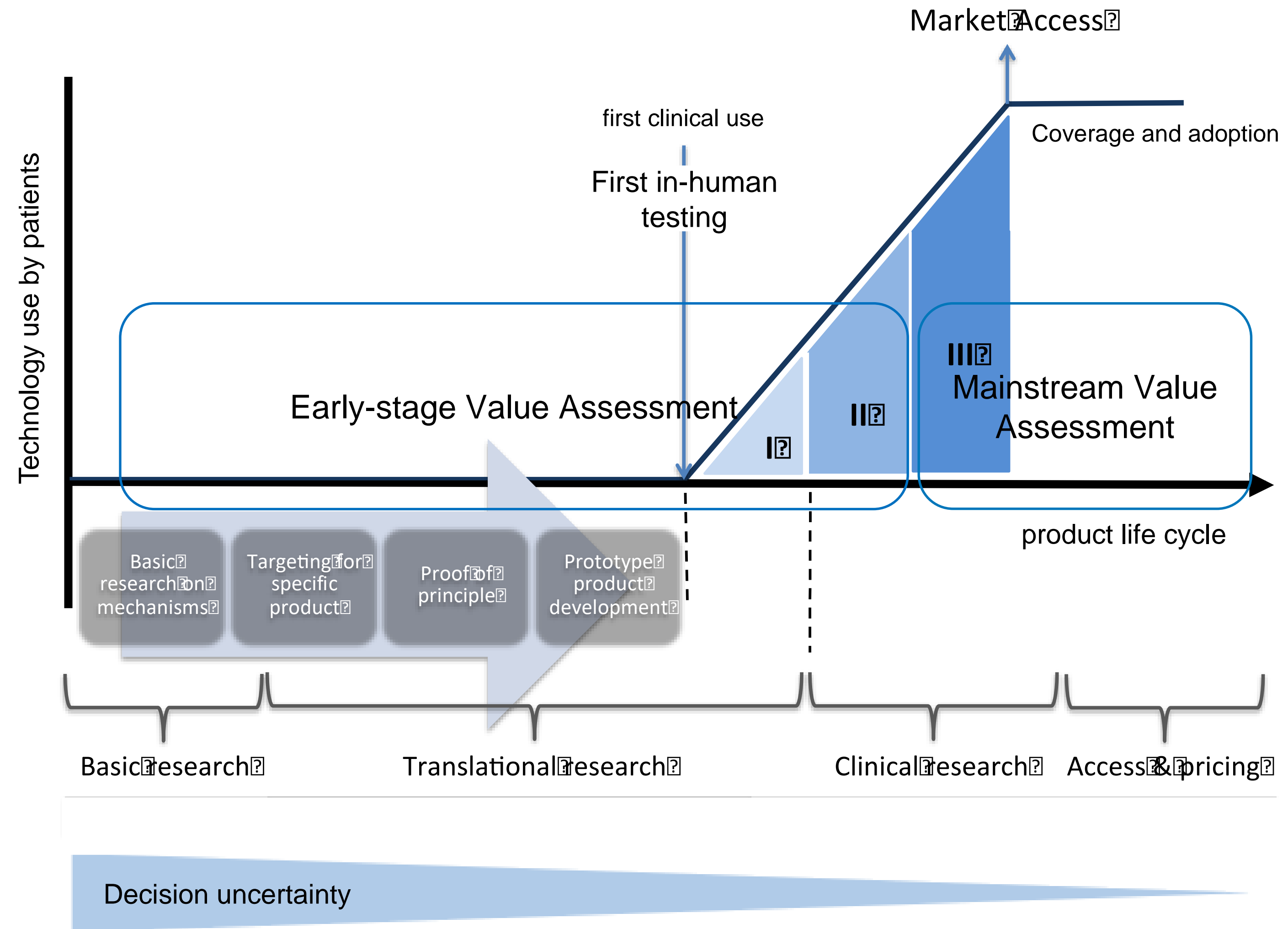
- Define the target population and clinical setting intended for use of the biomarker or screening test.
- Define subject inclusion and/or exclusion criteria and process for enrollment.
- Define the setting for specimen collection.
- Ensure adequate generality in the population studied.

Outcome

- Define the outcome of interest (early detection versus mortality reduction).
- Specify procedures for ascertaining and measuring the outcome.

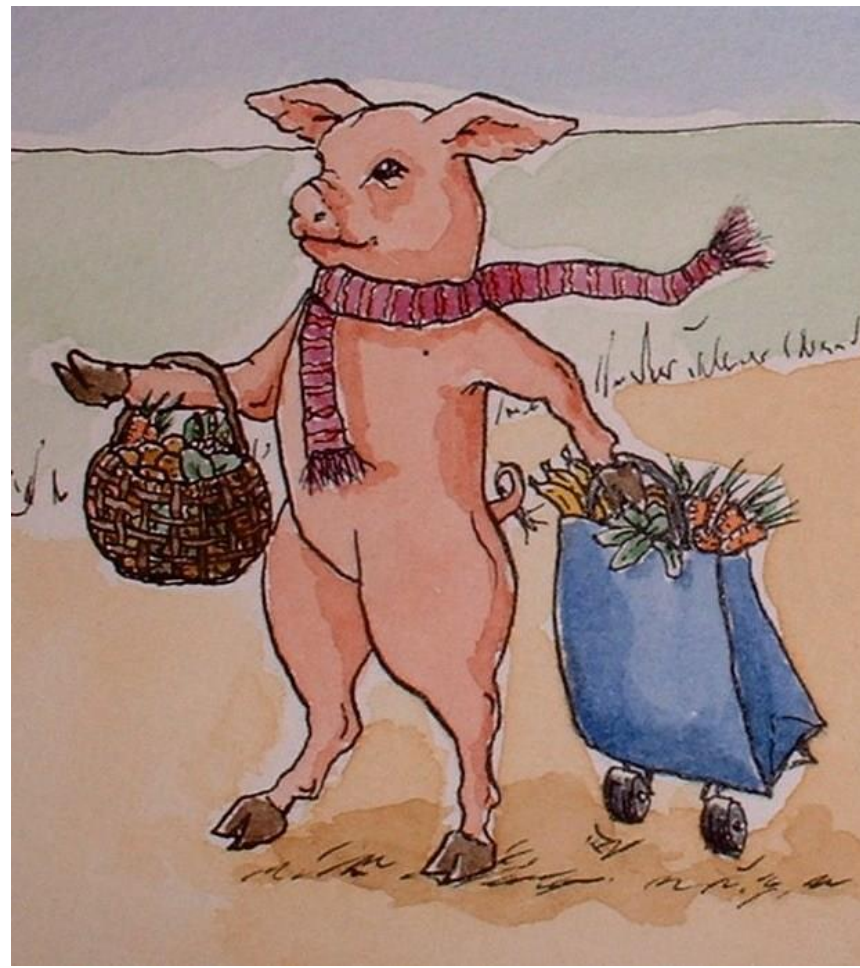
What comparator is acceptable?





IJzerman & Steuten, Applied Health Econ Health Policy 2011



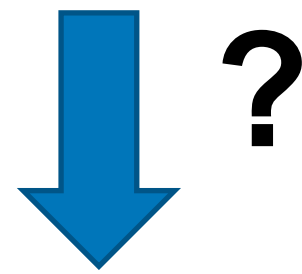


**This Little Piggy
Went to Market**

Guidelines USPSTF

Regulatory FDA

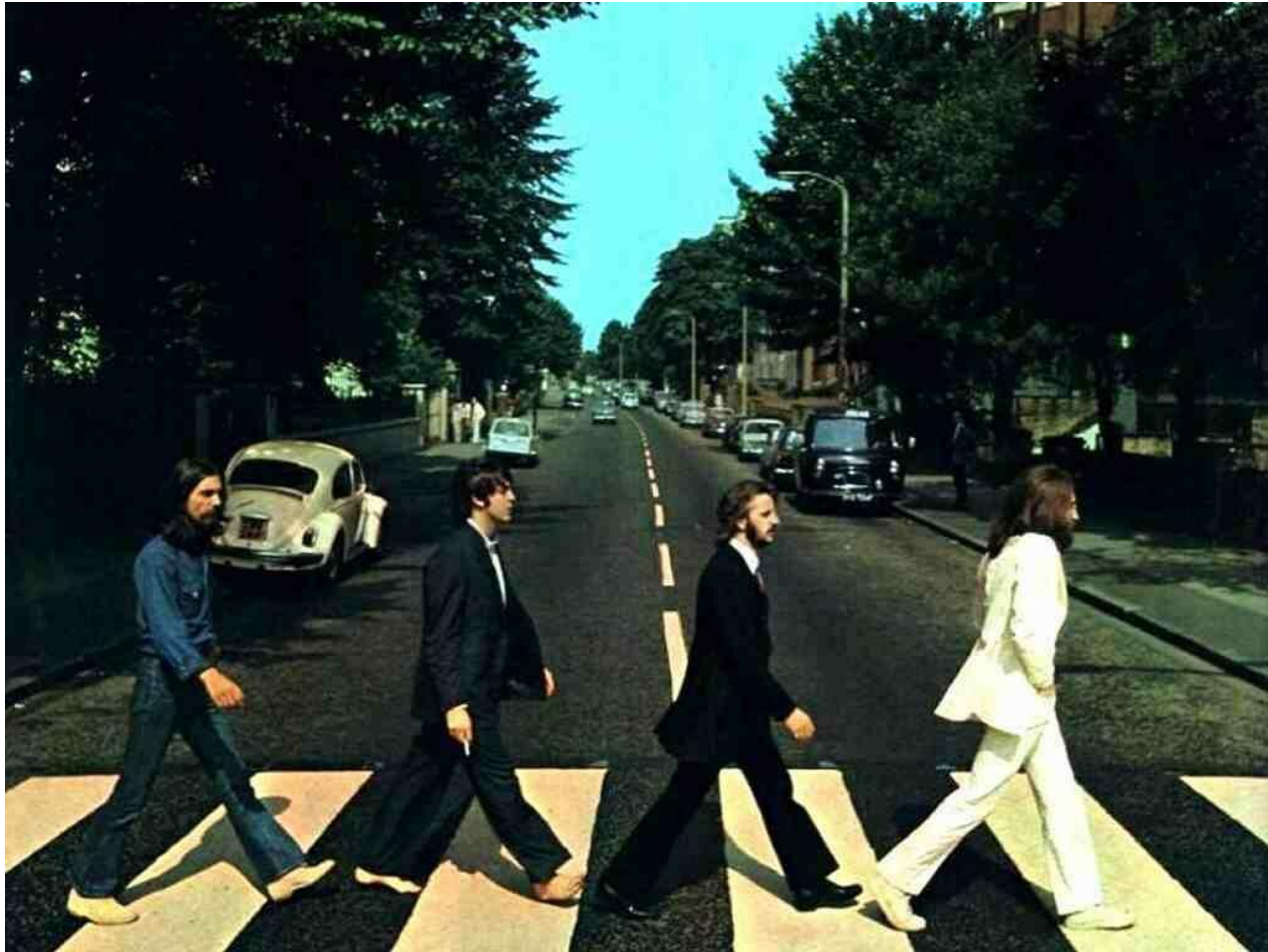
Payors CMS



**The Yellow
Brick Road
to Market**



The Long and Winding Road



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, AGAF¹; Carlo Senore, MD, MSc²; Jack S. Mandel, PhD, MPH³;
James E. Allison, MD, FACP, AGAF⁴; Wendy S. Atkin, MPH, PhD⁵; Robert Benamouzig, MD, PhD⁶; Patrick M. M. Bossuyt, PhD⁷;
Mahinda De Silva, MB, BS, FRACP⁸; Lydia Guittet, MD, PhD⁹; Stephen P. Halloran, MBE, FRCPath¹⁰; Ulrike Haug, PhD¹¹;
Geir Hoff, MB, ChB, PhD¹²; Steven H. Itzkowitz, MD, FACP, FACG, AGAF¹³; Marcis Leja, MD, MBA, PhD, AGAF¹⁴;
Bernard Levin, MB, BCh, FACP¹⁵; Gerrit A. Meijer, MD, PhD¹⁶; Colm A. O'Morain, MD¹⁷; Susan Parry, MbChB, FRACP¹⁸;
Linda Rabeneck, MD, MPH, FRCPC¹⁹; Paul Rozen, MD^{20†}; Hiroshi Saito, MD, PhD²¹; Robert E. Schoen, MD, MPH²²;
Helen E. Seaman, BSc, PhD²³; Robert J. C. Steele, MD, FRCS²⁴; Joseph J. Y. Sung, MD, PhD²⁵; and Sidney J. Winawer, MD²⁶

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 previous meetings and webinar.



Guiding Principles and Planning

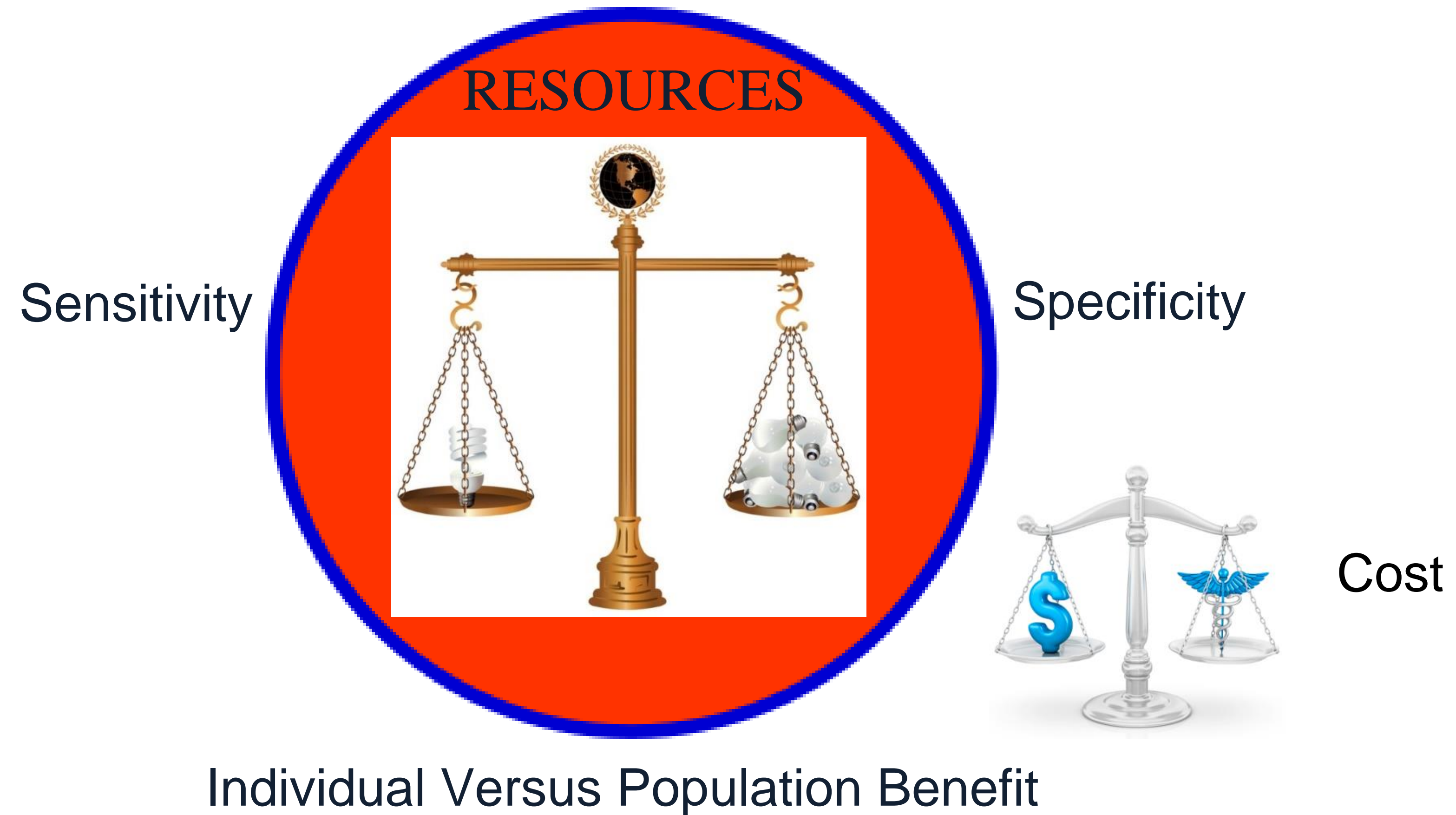
To revise the guiding principles for test evaluation, we established a consensus process based on the Glaser and Delphi approaches but adapted so as to be undertaken by a combination of webinars and voting via the internet due to the constraints of the COVID-19 pandemic. The *membership* was chosen from experts (including those available from the original report) because of their knowledge or experience in practice or research relevant to screening for CRC. Consensus of 80% agreement was desired (5 grade scale)

**Tested using a modified Delphi approach in 47 experts from WEO.
So far, 4 rounds have been completed.**



Colorectal Cancer Screening

How Good Should a Test Be?



2022 New Tests Comparison Consensus Process Proposed Revised Recommendations

1. Screening for colorectal cancer aims to reduce CRC mortality and/or incidence through detection of readily-treatable cancer and relevant precursor lesions. 94%; 95.7%
2. The screening test is just one step in a coordinated multistep process that includes initial and repeated participation by the intended-use population, quality-assured testing, diagnostic follow-up, treatment, and referral to high-risk surveillance programs when appropriate together with monitoring of key endpoints, Goals for each step in the process should be defined and agreed by providers. 87%; 93.5%
3. Randomized controlled trials (RCTs) with CRC mortality and/or incidence reduction as the primary outcome have set the expectations for performance of new tests. Effectiveness of a new test may be predicted when compared to a standard (i.e comparator or index test) where the comparator's effectiveness has been demonstrated. 78.2-89%; 87.0%

* Rounds 1 & 2; Round 3

4. The performance of a new non-invasive screening test can be assessed in parallel to an existing non-invasive screening test of proven effectiveness at any step in the screening process from population engagement to key outcomes. Intermediate endpoints known to reliably and consistently predict potential for reducing colorectal cancer mortality and/or incidence should be used to compare a new test with existing screening tests. Such endpoints include estimates of test accuracy (primarily true- and false-positive rates and sensitivity and specificity). 89%; 89.1%

5. In two-step screening based on first performing a non-invasive test followed by colonoscopy if positive, the test should identify participants with an increased likelihood of CRC or relevant precursor lesions. 80.9-91%; 89.1%

6. The precursor lesions currently considered to be of sufficiently high risk to be important to detect are advanced adenomas and advanced serrated lesions. More research is needed to clarify how best to characterize those lesions that are most important to detect and remove. 93%; 89.1%

7. A non-invasive screening test with an adjustable positivity threshold (or algorithm) enables choice of test accuracy parameters (sensitivity, specificity) and test positivity rate that best achieve desired goals of a screening program. Regulatory approval processes for a new test should consider capacity for adjustment of the positivity-threshold by screening-program providers or policy-makers, when alternative performance characteristics are better suited to meet program goals. 74.5-91%; 84.8%

8. Evaluation of a new test should follow a four-phase (sequential) evaluation. This would start with limited-scale cohort or case-control studies in populations with and without neoplasia, possibly enriched for the neoplastic outcomes of interest. Initial estimates of clinical performance characteristics and positivity criteria will be obtained in phase 1 and 2 studies. If results suggest that the test might achieve the desired standard, evaluation should proceed to screening pathway evaluation requiring larger *intended-use* screening populations (Phases 3 and 4). The latter studies should be prospective and will identify the most suitable criteria for test-positivity, amongst other important outcomes. 89%; 87%

9. The desired accuracy considered to be sufficient to proceed to intended-use population evaluation in phase 3, will be subject to a range of considerations which vary between jurisdictions. Nonetheless, it is considered ideal if phase 1 and 2 studies demonstrate that test accuracy of a new test is at least comparable to that of non-invasive tests in existing public-health screening programs. 85% then 76.1%; 100%

10. Non-invasive tests targeting new biomarkers might detect lesions that differ in their natural history from those detected by established tests. Cancers detected by a new biomarker might be more or less responsive to treatment. Consequently the benefit from treatment might not be the same. Precursor lesions detected by a new biomarker might identify a different risk profile (genotype or phenotype). Exploring concordance between the new test and a comparator test determines the degree to which lesion detection varies. 74.5%-89%; 84.8%

11. If a non-invasive test is to be widely used in screening programs and be fully supported in guidelines, its application at key points along the multi-step screening pathway should be assessed in the intended-use population. In addition to sensitivity, specificity, measures would include acceptability to invitees, test failure rates, colonoscopy workloads, and eventually 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. Alternatively modelling studies mimicking such RCTs and based on high quality observational data will also be informative. 91%; 84.8%

12. Prior to undertaking assessments of the clinical value of a new test in an intended-use population, it is essential that the analytical performance characteristics of a test are documented according to relevant standards, such as those of the Clinical and Laboratory Standards Institute (CLSI). This information should be provided by the test manufacturer in their instructions for use. Ideally, evaluation of analytical characteristics should be performed using recommended protocols such as are described by the CLSI. Researchers undertaking development of a new test should follow such protocols and undertake required verification processes for test development to ensure that the prescribed standards are attained or surpassed. The international standard ISO 15189 specifies the requirements for competence and quality that apply to medical laboratories. Ideally, the laboratory sites at which test are performed should be also accredited to applicable standards, such as specified in ISO 15189.

85%; 87%



Bert Vogelstein is Co-director
of the Ludwig Center

**... if the current trends continue,
the number of cancer cases diagnosed
annually by 2050 is likely to double as
a result of population aging. So if we
as a society hope to head off the
coming storm, we better get more
serious about cancer prevention soon.**

**“ If people are constantly falling off a cliff, you could place
ambulances under the cliff or build a fence on top of the
cliff. We are placing too many ambulances under the cliff ” –
Denis Burkitt**

