WEO

The voice of world endoscopy

An efficient strategy for evaluating new non-invasive screening tests for colorectal cancer: The guiding principles

Authors: Bresalier* R.S., Senore*, C., Young*, G.P., Allison, J., Benamouzig, R, Benton, S., Bossuyt, P.M., Caro, L., Carvalho, B., Chiu, H.M., Coupe, V.M.H., de Klaver, W., de Klerk, C.M., Dekker, E., Dolwani, S., Fraser, C.G., Grady, W.M., Guittet, L., Gupta, S., Halloran, S.P., Haug, U., Hoff, G., Itzkowitz, S.H., Kortlever, T.L., Koulauzidis, A., Ladabaum, U., Lauby-Secretan, B., Leja, M., Levin, B., Levin, T.R., Macrae, F., Meijer, G.A., Melson, J., O'Morain, C., Parry, S., Rabeneck, L., Ransohoff, D.F., Saenz, R., Saito, H., Sanduleanu, S., Schoen, R.E., Selby, K., Singh, H., Steele, R.J.C., Sung, J.J.Y., Symonds, E., Winawer, S.J. (Members of the WEO CRC Screening New Test Evaluation Expert Working Group)

Gut in Revision

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[1]; Carlo Senore, MD, MSc[2]; Jack S. Mandel, PhD, MPH[3]; James E. Allison, MD, FACP, AGAF[4]; Wendy S. Atkin, MPH, PhD[5]; Robert Benamouzig, MD, PhD[6]; Patrick M. M. Bossuyt, PhD[7]; Mahinda De Silva, MB, BS, FRACP[8]; Lydia Guittet, MD, PhD[9]; Stephen P. Halloran, MBE, FRCPath[10]; Ulrike Haug, PhD[11]; Geir Hoff, MB, ChB, PhD[12]; Steven H. Itzkowitz, MD, FACP, FACG, AGAF[13]; Marcis Leja, MD, MBA, PhD, AGAF[14]; Bernard Levin, MB, BCh, FACP[15]; Gerrit A. Meijer, MD, PhD[16]; Colm A. O’Morain, MD[17]; Susan Parry, MbCHB, FRACP18[18]; Linda Rabeneck, MD, MPH, FRCPC[19]; Paul Rozen, MD[20†]; Hiroshi Saito, MD, PhD[21]; Robert E. Schoen, MD, MPH[22]; Helen E. Seaman, BSc, PhD[23]; Robert J. C. Steele, MD, FRCS[24]; Joseph J. Y. Sung, MD, PhD[25]; and Sidney J. Winawer, MD[26]

Cancer 2016; 122(6):826-39

What has transpired?

- New developments in biomarker technologies.
- Widespread implementation of organized population screening that makes test evaluation difficult in intended-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.

- **Glaser and Delphi** approaches adapted to be undertaken by a combination of webinars and voting via virtual platforms due to the constraints of the COVID-19 pandemic (in-person discussion during DDW 2022)
- The membership consisted of experts (gastroenterologists, endoscopists, gastrointestinal surgeons, public health physicians, epidemiologists, clinical biochemists and tumor biologists) with knowledge or experience in practice or research relevant to screening for CRC. **Forty-seven experts** were involved.
- A series of specific questions (each of which was a draft principle to be critiqued) was initially expanded from the original eight to ten and then, after the first consensus round of voting, further increased to 12. The **12 principles** were progressively redrafted in response to specific feedback: webinars, conference seminars addressing specific issues and semi-structured discussions were held, and members voted and commented on each principle using a spreadsheet. After four rounds of voting, the consensus goal of **>80% agreement** (agree or strongly agree on a 5-point scale) was achieved for all 12 principles.
- The **explanatory text** for each principle was developed from the feedback received during the consensus process and from the extensive comments received during the consultation of experts and industry representatives. Multiple drafts of the explanatory text were circulated to the expert panel over a period of six months, and feedback has been incorporated into the final manuscript.

Colorectal cancer is a global disease, and a “one size fits all” approach to CRC screening may not be relevant. Guiding Principles, however, are necessary and should be universal. The epidemiology of CRC will undoubtedly change over time which may alter the composition of intended use populations. **We present a framework that allows a dynamic process that has broad application.** This process is not bound by any one specific test.

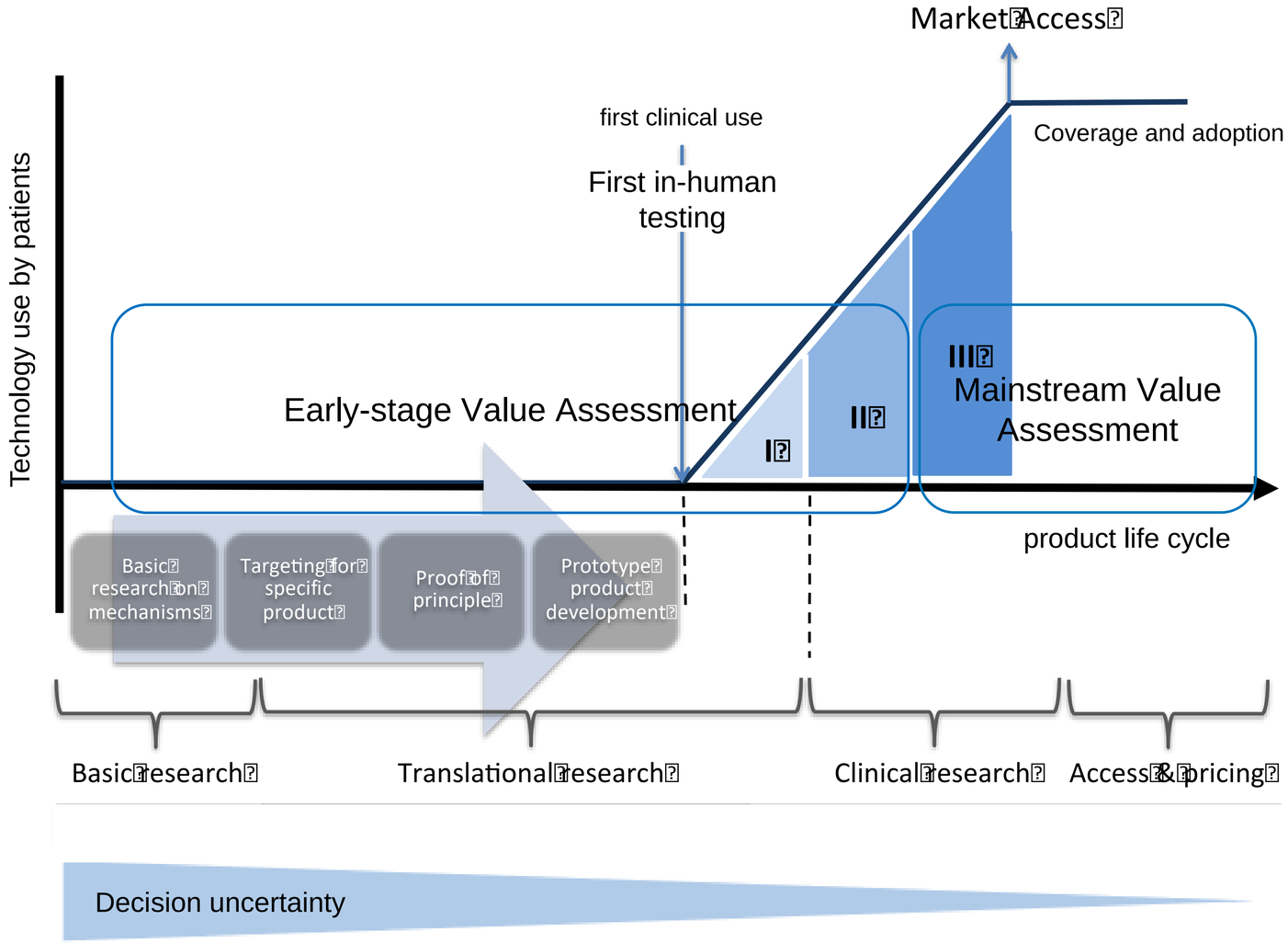
Colorectal Cancer Screening
How Good Should a Test Be?

Sensitivity

Specificity

Cost

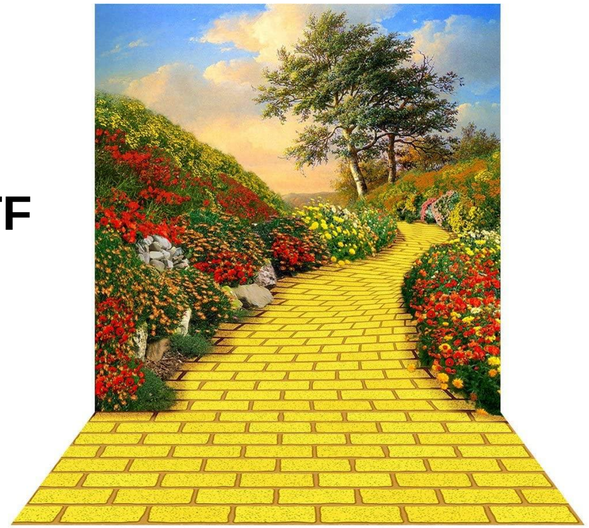
Individual Versus Population Benefit



IJzerman & Steuten, Applied Health Econ Health Policy 2011

This Little Piggy Went to Market

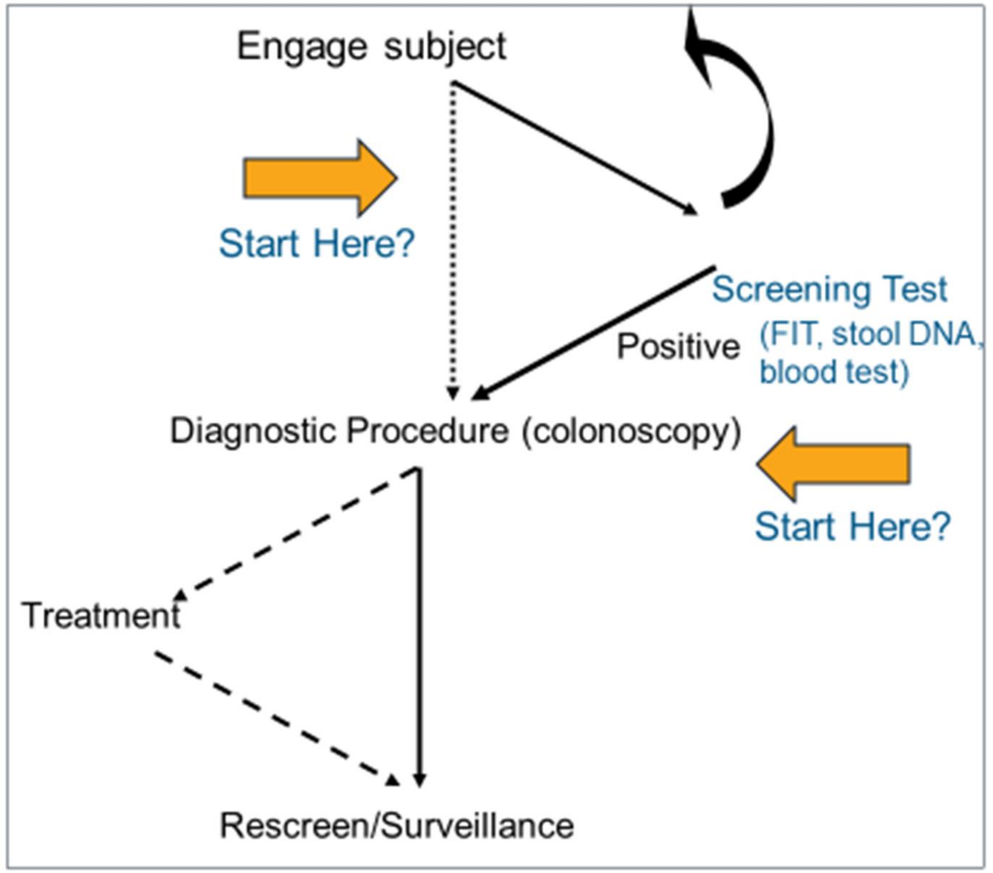
?



The Yellow Brick Road to Market

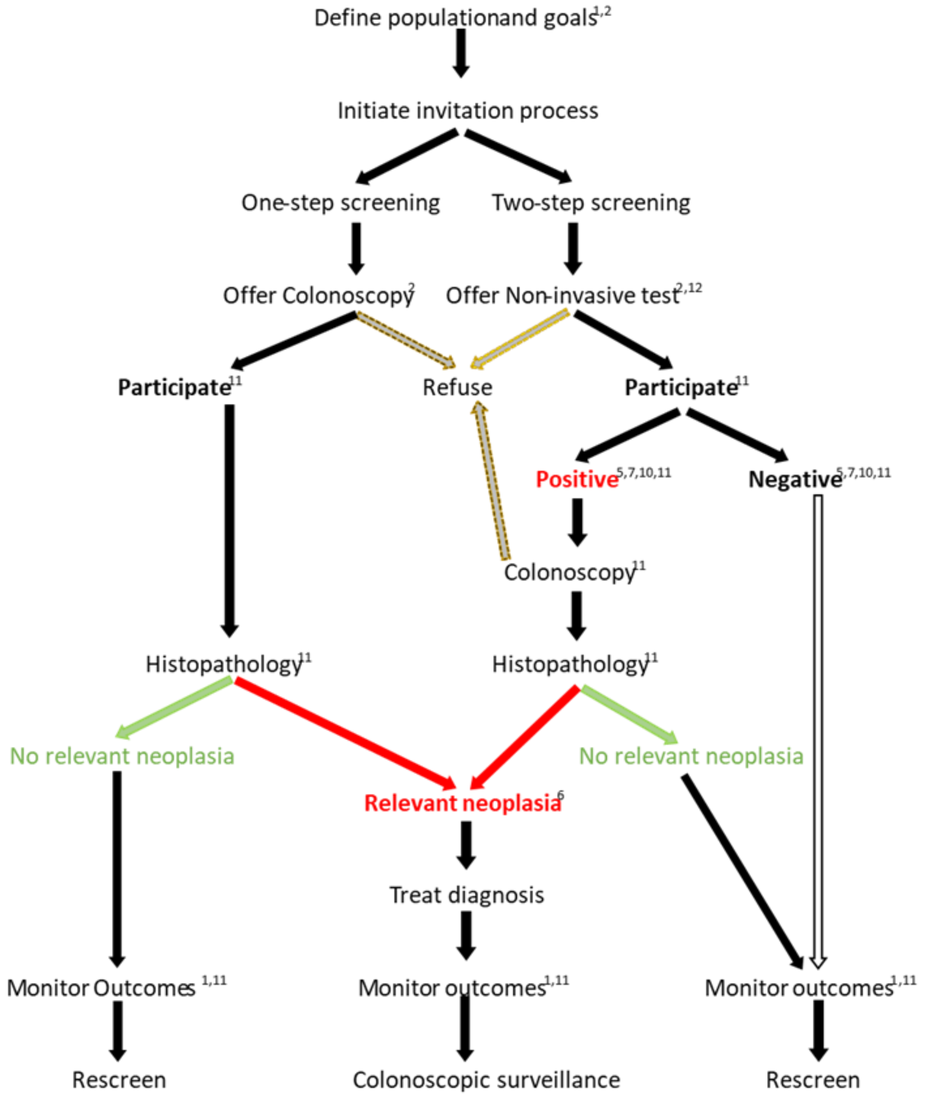
Guidelines	**USPSTF**
Regulatory	**FDA**
Payors	**CMS**

One-Step Versus Two-Step Screening



aga

Multistep screening pathway characteristic of organized screening programs and demonstrating one- and two-step strategies



Topics Addressed in Each of the Principles Established by the Consensus Process

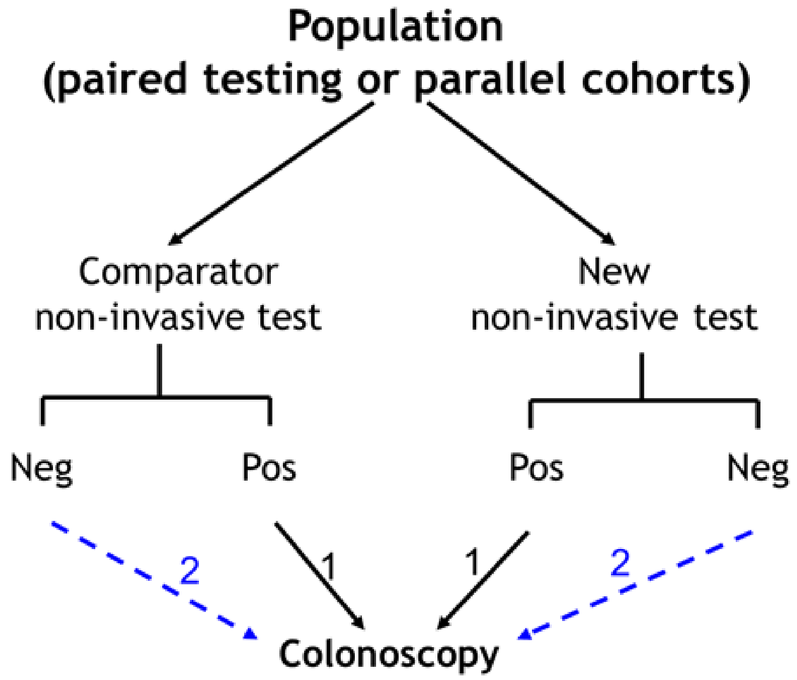
Principle Number	Topic
1	Desired outcomes of CRC screening
2	Screening is a multi-step process
3	A screening test identifies individuals with an increased likelihood of CRC and/or advanced precursor lesions
4	Nature of precursor lesions most important to detect
5	New biomarkers might detect lesions with a different natural history
6	Outcomes to be estimated in a screening population
7	Expectations of a new non-invasive test
8	An adjustable test positivity threshold accommodates different program goals
9	Predicting value by paired comparison to a proven non-invasive test
10	Evaluation proceeds through increasingly complex phases
11	Accuracy required for evaluation in a screening population
12	Analytic specifications, standards, and performance

A rigorous and efficient four-phased approach is proposed

- Commencing with small studies to assess the test's ability to discriminate between CRC and non-cancer states (Phase 1)
- Followed by prospective estimation of accuracy across the continuum of neoplastic lesions in neoplasia-enriched populations (Phase 2).
- If these phases show promise, a provisional test-positivity threshold is set before evaluation in a typical screening population.
- Phase 3 prospective studies in a single screening round determine intention-to-screen program outcomes.
- Phase 4 studies involve evaluation over repeated screening rounds with monitoring for missed lesions.

Phase 3 and 4 findings will provide the real-world data required to model test impact on CRC mortality and incidence.

Diagrammatic outline of trial design appropriate for comparing non-invasive tests in the initial phases of test evaluation



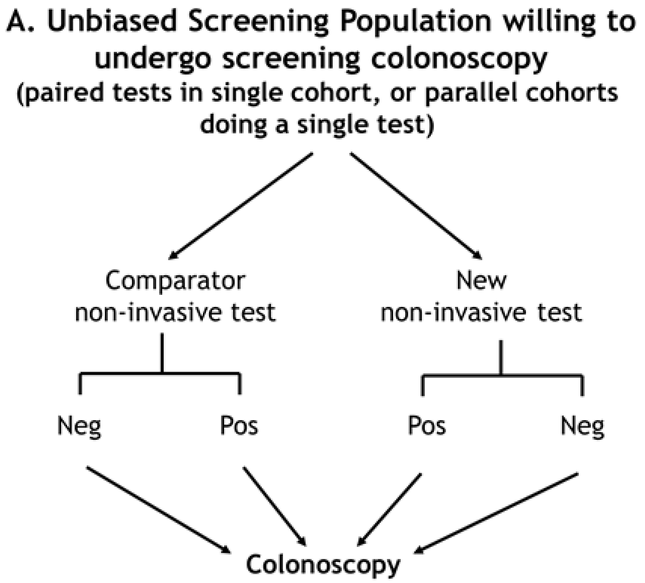
1 For comparing true- and false-positive fractions.
2 For comparing sensitivity and specificity (depending on biases due to population selection).

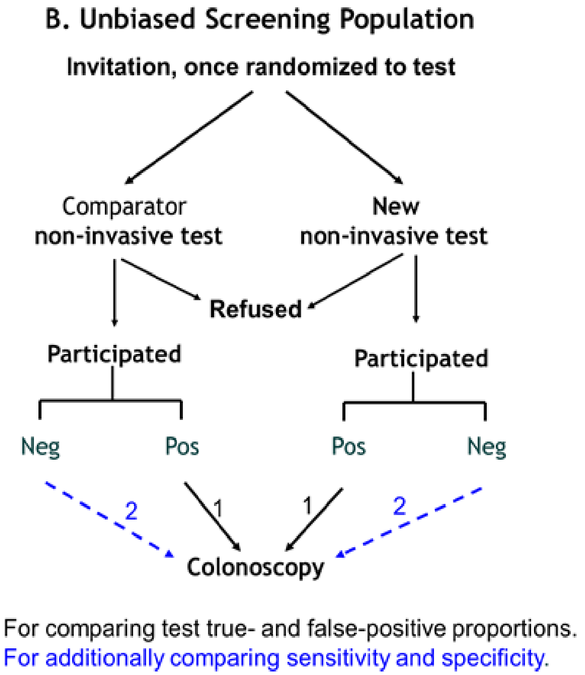
Paired testing is conducted in a single cohort where an individual does both the new and the comparator test, whereas parallel testing is where study participants are randomized to one or the other test

Phased (sequential) stepwise evaluation is an efficient way first to establish the potential value of a new test and then to subsequently gather the evidence that will lead to its acceptance by professionals, healthcare providers, and regulatory bodies

Phase	Goal(s)	Context	Approach and measures	Hurdle for progression
1	**Main:** Differentiates between CRC and non-neoplastic states?	Prescreening cohorts – limited	Distribution of test results in cohorts with and without CRC	• Test result must differ significantly in cancer cases.
2	**Main:** Detects early cancer and precursor lesions? **Others:** Initial positivity threshold? Accuracy relative to comparator? Causes of false-positives.	Prescreening cohorts - extensive	Distribution of test results in cohorts with CRC relevant precursor lesions, other colorectal diagnoses and no disease. Parallel or paired testing of new and comparator tests will be informative.	• Preliminary (although biased) estimates of accuracy are shown to be promising. • ROC analysis identifies a suitable positivity threshold.
3	**Main:** Test accuracy in a typical screening evaluation? Test acceptance? **Others:** Test failure rate? Other variables for modelling effectiveness and cost-effectiveness.	Screening populations – single round	Apply test prospectively to a typical unbiased intended-use population. Choose study design appropriate to program goal and jurisdictional context: e.g., colonoscope all for estimating test accuracy, parallel testing for comparing non-invasive tests and intention-to-screen outcomes.	• A significant improvement in some aspect of screening. • Non-inferior in accuracy to a comparator test, OR • Accuracy likely delivers benefit. • Feasible colonoscopy workload. • Modeled effectiveness and cost-effectiveness are satisfactory.
4	**Main:** Missed lesions or adverse events? **Others:** Participation rates over time and re-test intervals?	Screening population – multiple rounds	Apply the test prospectively to an intended-use screening population over multiple rounds, with careful monitoring of population program outcomes.	

Study design frameworks applicable to Phase 3 studies





1 For comparing test true- and false-positive proportions.
2 For additionally comparing sensitivity and specificity.

A: design appropriate to determine test accuracy where all cases undergo colonoscopy, but intention-to-screen outcomes cannot be ascertained (comparison of a comparator with the new test can be paired in a single cohort or parallel in separate cohorts). B: design appropriate for estimating intention-to-screen outcomes and where the accuracy of the new test can be compared to that of a non-invasive comparator either when colonoscoping only test-positive individuals (compare true- and false-positive fractions) or all participants (sensitivity and specificity).

Next Steps

- Revision and response to reviewers completed. Final product to be submitted within 2 weeks (Gut)

- Plan a series of follow-up papers addressing expanding key elements and controversies (to be published as a special addition to a journal (preliminary agreement in place)

❑ Trial design

❑ What is the appropriate target endpoint

❑ How good is “good enough”

❑ Flexibility and expectations from existing trials

❑ Power and statistical considerations

❑ Modeling and use of surrogates

❑ The role of academic/industrial partnerships

❑ Regulatory challenges

WEO

World Endoscopy Organization