

The authors would like to take this opportunity to thank the reviewers for their time and attention to this manuscript. Our responses to your comments are set forth below.

Reviewer #00068256

The present manuscript by Berger et al described the possibility and advantages of Multi-target stool DNA testing over other recommended screening options for CRC, using USPSTF's modeling data. They concluded that mt-sDNA at 3 year intervals is a reasonable approach for general CRC screening. Indeed, the value of Multi-target stool DNA testing has been confirmed by many related studies, several factors also should be discussed in this paper, including test price and availability in less developed regions or countries (instruments et al).

Authors' response: A section indicating the current status of test cost and availability outside the U.S. has been added. A full discussion on the challenges of introducing a new technology world-wide are outside the scope of this topical review. However, we look forward to providing an update in the future.

The following has been added to the manuscript (page 3 of the manuscript - highlighted):

"The test is primarily being used in the United States and has been approved by United States Food and Drug Administration for average risk CRC^[6] and is reimbursed by the U.S. Centers for Medicare and Medicaid Services once every three years^[7]. It was awarded a unique Clinical Procedural Terminology (CPT) code of 81528 by the American Medical Association. Cologuard[®] (multitarget sDNA) has been CE marked for use in Europe, though there is limited availability to date through laboratories England and Dubai. Additional studies to determine local efficacy in a number of Asian countries are under discussion and studies are ongoing in Italy, the United Kingdom and the Netherlands. The test cost is U.S. \$649, which includes a U.S. nationwide patient navigation/compliance system supporting over 70 languages."

Reviewer #00057665

This is a timely review of the use of multi-target DNA for colorectal cancer screening. Anyway, I am puzzled by the use of the term "efectiveness". Shouldn't "eficacy" be used?, given that most studies were prospective ones and they do not represent "real world data".

Authors' response: Agreed, we have corrected the language to reflect "efficacy."

Reviewer #01490490

1) The authors do not introduce the concept of opportunistic vs organized (programmatic) screening, which is crucial to understand their recommendations. In the setting of opportunistic screening, as it is the case in USA, screening with mt-sDNA testing is clearly an alternative as it has been demonstrated its high sensitivity for detecting advanced colorectal neoplasia. However, in the setting of population-based organized screening is not a real practice alternative at the present time, because there is lacking evidence to support it (i.e. there is no data on screening uptake and costs). So, it looks like this review has been written for countries in which only have opportunistic CRC screening;

Authors' response: Multitarget stool DNA is being used in the United States in both opportunistic and invitational settings. The structure of the test, which require no preparation and a random stool sample, lends itself to population-based use while the size of the specimen container and cost might limit its use in that manner. Countries with limited colonoscopy capacity to follow up on positive tests will also be challenged by the lower specificity compared to FIT, albeit at the loss of a significant percentage of curable colorectal cancers, high grade dysplasias and almost any detection of sessile serrated adenomas due to the lower sensitivity of FIT as an alternative.

The following has been added to the manuscript (page 4 of the manuscript - highlighted):

"Multitarget stool DNA is being used in the United States in both opportunistic and local invitational settings. The structure of the test which includes at-home stool sample collection and requires no change in diet, medication or any preparation, lends itself to population-based use while the size of the specimen container, allowable three-day transit time back to the laboratory, and cost might mitigate use in that manner, especially in low resource countries and those with predominantly rural populations. Countries with limited colonoscopy capacity to follow up on positive tests will also be challenged by the lower specificity of mt-sDNA compared to FIT, albeit at the loss of a significant percentage of curable colorectal cancers, high grade dysplasia and almost any detection of sessile serrated adenomas due to the lower sensitivity of FIT as an alternative."

2) This review is only focused on the benefits of mt-sDNA testing, but does not give any insights about the test limitations. First, the test costs \$599, according to CEO Conroy, which makes it impossible for using in population-base screening nowadays.

Authors' response: As noted in our response to Reviewer #00068256 we have included information regarding the availability and cost of the test by adding the following paragraph to the description of the test (page 3 of the manuscript - highlighted):

"The test is primarily being used in the United States and has been approved by United States Food and Drug Administration (FDA) for average risk CRC and is reimbursed by the U.S. Centers for Medicare and Medicaid Services (CMS) once every three years. It was awarded a unique Clinical Procedural Terminology (CPT) code of 81528 by the American Medical Association. Cologuard® (mt-sDNA) has been CE marked for use in Europe, though there is limited availability to date. Cologuard (U.S. \$649) is available through laboratories in England and Dubai. Additional studies to determine local efficacy in a number of Asian countries are under consideration."

Second, in one round, sensitivity for advanced adenoma is lower than 50% and specificity is lower than 90% (high number of false positives that leads to perform a high number of unnecessary colonoscopies).

Authors' response: Countries with limited colonoscopy capacity to follow up on positive tests may be challenged by somewhat lower single event "clean colon" specificity (90%) of mt-sDNA compared to FIT. However, mt-sDNA every 3 years results in similar or fewer negative colonoscopic follow-up examinations than the compounded effect of annual or biannual FIT at 95% specificity. Further, non-annual use of FIT may fail to identify a significant percentage of curable colorectal cancers, high grade dysplasia and almost any detection of sessile serrated adenomas due to the lower sensitivity of FIT.

Third, in the Deep-C study (Imperial et al NEJM 2015), the most relevant study on mt-sDNA, including almost 10.000 individuals, about 6% of patients who had the stool DNA test, for whatever reason the samples couldn't be analyzed. In the context of the baseline prevalence of colon cancer in this population, that means that they lost the results for 689 patients, and 4 cancers could have been missed just because the test could not be done. So, there are important caveats about mt-sDNA that the authors should comment in this review. I suggest to include a specific paragraph explaining the limitations of this test.

Authors' response: Regarding the 6% of Deep-C study patients whose initial sample were not analyzable; in daily practice we simply get a second specimen. In the Deep-C study situation there was not sufficient time to get a second specimen returned as the patient generally had already gone to colonoscopy. It is important to note that that the most common reason that patients in the Deep-C study did not have results was that they failed to appear for colonoscopy. While it is possible that there were additional cancers in that subgroup, there is no evidence that there would be something unique about these possible cancers

that would change the overall performance difference seen between FIT and mt-sDNA in the other 65 cancers identified by colonoscopy in the study.

By way of comparison, in daily practice approximately 10% of colonoscopies fail because of a variety of issues, predominantly poor prep, and have to be repeated and fewer than 50% of patients provided with FIT tests, even in highly controlled health system, return them for testing. Currently, 71% of patients receiving a mt-sDNA kit return it for screening within 60 days. This is higher compliance in the U.S. than with any of the other screening tests at the current time. There is room for improvement with all the CRC screening tests.

Regarding the suggestion to include a specific paragraph explaining the limitations of this test; in the United States specifically and many other countries in general, colonoscopy is a well accepted colorectal cancer screening test. The purpose of a screening test is to screen for clinically relevant, treatable disease. We believe that the trade off of 10% colonoscopies with no findings against 94% detection of stage I and II colorectal cancer and 69% of high grade dysplasia is worth the tradeoff, especially in the United States where colonoscopy itself is a screening test. Further, in countries that use annual FIT or FOBT, the cumulative three-year specificity failure ($5\% \times 3 = 15\%$) is similar to the specificity of mt-sDNA done once every three years. This can be seen in the modeling exercise for 2 and 3 year FOBT/FIT testing included already in the paper.

To ensure that the discussion on test limitations is full we have amended that discussion further to include the following language from the FDA Summary of Safety and Effectiveness Data (added to page 13 - highlighted):

"Like all tests, mt-sDNA may be associated with false positive results and false negative results, wherein advanced colorectal neoplasm is not identified on a single screening event. Colonoscopy may be associated, though rarely, with significant adverse events.^[23]"

3) Based on hypothetical cost-effectiveness and modeling analysis the authors recommend to repeat mt-sDNA at three-year intervals in the setting of average-risk screening. However, there are no studies supporting this assumption in real-world practice, a necessary condition to establish testing intervals for CRC screening. The authors should make clear that prospective studies are needed to clarify which is the best time interval for mt-sDNA testing in the setting of CRC screening

Authors' response: The reviewer asks that we make clear that prospective studies are needed to clarify the optimal time interval for mt-sDNA screening. We respectfully disagree with that contention given the challenges of designing such a study when all the covariates are account for (previous screening of subjects, a priori risk, compliance etc.). We have had extensive discussion with GI leaders on this including both US (Ann Zauber) and European (Ernst Kuipers) and CISNET modeling group investigators. The current state of the art for establishing the initial test interval for a new test is modeling (U.S. Agency for HealthCare Research and Quality, due to the size, complexity and length of studies required for a confident analysis of multiple intervals (AHRQ CER 52, added to references). Only low sensitive guaiac based FOBT and flexible sigmoidoscopy have intervals supported by RCT's and those intervals may not even be optimal as compliance with annual biannual screening, which is recommended, is poor. The modeling is well described already in the paper and is further supported by a recently published study.

To ensure that the discussion on test interval is fully addressed we have amended that the manuscript to include the following language (added to page 6 - highlighted):

"Establishing an initial inter-test interval using vetted, well designed and calibrated CRC screening models is currently recommended for new tests.²⁶ Prospective longitudinal studies are large, complex, lengthy and expensive. Currently only use of low sensitivity guaiac FOBT and

WJGO Manuscript NO 23899

January 26, 2016

Berger

flexible sigmoidoscopy are supported by prospective randomized control trials. Models allow for virtual prospective studies to be done on very large cohorts of virtual people and provide comparative performance of multiple tests simultaneously. Such modeling and its limitations are described below."

We respectfully submit these edits and hope that they have strengthened the manuscript and addressed the reviewers' concerns and comments. We look forward to your response and are open to further considerations, if needed.

Best regards,

Barry M. Berger, MD

Corresponding Author