Stool Testing for Colorectal Cancer Screening

Douglas J. Robertson¹

Thomas F. Imperiale²

¹ Department of Veterans Affairs Medical Center, White River Junction, Vermont, and Dartmouth Medical School and Dartmouth Institute, Hanover, New Hampshire

² Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana; and Regenstrief Institute, Inc, Center of Innovation, Roudebush Veterans Affairs Medical Center, Indianapolis, Indiana

Abstract

Colorectal cancer (CRC) screening has been shown to reduce CRC incidence and mortality and is widely recommended. However, despite the demonstrated benefits of screening and ongoing efforts to improve screening rates, a large percentage of the population remains unscreened. Noninvasive stool based tests offer great opportunity to enhance screening uptake. The evidence supporting the use of both fecal immunochemical testing (FIT) and stool DNA (sDNA) has been growing rapidly and both tests are now commercially available for use. Other stool biomarkers (eg, RNA and protein based) are also actively under study both for use independently and as adjuncts to the currently available tests. This mini review provides current, state of the art knowledge about noninvasive stool based screening. It includes a more detailed examination of those tests currently in use (ie, FIT and sDNA) but also provides an overview of stool testing options under development (ie, protein and RNA).

Keywords: Biomarkers; Colorectal Cancer Screening; Fecal Immunochemical Test; Stool DNA

This is the author's manuscript of the article published in final edited form as:

Stool Testing for Colorectal Cancer Screening

Colorectal cancer (CRC) screening is available across the United States, Europe, and Asia. Unlike screening tests for breast and prostate cancer, there is strong evidence that CRC screening reduces disease-specific morbidity and mortality. For fecal occult blood testing and flexible sigmoidoscopy, several large randomized trials show a mortality benefit, whereas observational studies show a mortality benefit for colonoscopy relative to usual care.

Although CRC screening is widely available, the specifics of the programs vary across countries. Programmatic screening with fecal occult blood, flexible sigmoidoscopy, and colonoscopy has been successfully used in Canada, Europe and Asia. In the United States, screening is largely opportunistic and guidelines endorse a panel of options. Evidence suggests that colonoscopy is currently the dominant approach and so the role of the gastrointestinal specialist has largely been centered on providing high quality endoscopy. However, market forces may impact this practice. Gastrointestinal specialists likely will be increasingly relied upon to develop CRC screening programs for the populations they serve. In this capacity, knowledge about noninvasive stool based options will become increasingly important.

This mini-review provides current, state of the art knowledge about noninvasive stool based screening. The key advantages of noninvasive screening approaches are reviewed and the current state of evidence supporting such use summarized (Table 1). The review more closely details clinically available tests (ie, fecal immunochemical testing [FIT] and stool DNA [sDNA]) but also provides an overview of stool testing options under development (ie, protein and RNA).

Desirable Features of a Screening Test
Several features are desirable for a screening test or, when repeated testing over time is instituted, a screening program. Since screening involves testing asymptomatic people who are mostly healthy, a screening test must be low risk both directly (ie, the test itself must not cause harm) and indirectly (ie, downstream risks resulting from the need for subsequent testing). Related to this latter concept is the false-positive rate, which is a function of specificity. A test with high specificity and low false positive rate reduces risks of harm from both unnecessary testing and overdiagnosis.11 It also helps reduce the cost of screening, another desirable feature, particularly from a population-based perspective. Of course, the test should be able to detect the target lesion, making sensitivity of the test or program very important. For CRC screening, a noninvasive test as considered in this review should be able to detect curable-stage CRC, and if not on the first round of testing, then with the next round of testing, before a curable CRC becomes incurable. The degree to which a test should detect pre-cancerous lesions, most of which never progress to CRC, is debatable and beyond the scope of this review. Given the relatively long dwell time for neoplastic progression,12 a test with reasonable programmatic sensitivity may be adequate. Finally, a screening test should be widely available, affordable, and user-friendly, features required to optimize test uptake and adherence.

From a practical perspective, a “good” noninvasive test affords several advantages for both patients and health systems. Patient advantages include an opportunity to better match the screening test to individual preferences and the factors they value.13 There is evidence that offering individuals a choice of screening options improves adherence with subsequent testing.14 From a systems or societal perspective, availability of noninvasive testing may increase the reach of CRC screening programs to include at least some of the 35% of the US population who are
not current with CRC screening. To the extent that these noninvasive approaches are less costly than their structural counterparts, they can free resources for other worthwhile interventions.

**Quality of Evidence**

Evidence is lacking for which screening test and strategy are best and for the cumulative benefits and harms of screening. Optimal evidence would derive from a population-based randomized trial comparing all recommended screening strategies to either “no screening” or “usual care” with CRC mortality as the primary endpoint. Of course, logistical and ethical reasons preclude the feasibility of such a trial, resulting in the need to make clinical decisions using lower quality evidence.

It is clear that noninvasive screening with stool testing works and has durable benefits. Long-term follow-up of persons randomized to guaiac-based fecal occult blood test (gFOBT) or usual care demonstrated a 32% reduction in CRC mortality after 30 years of follow-up. However, gFOBT has limitations largely because of its reliance on detecting a more nonspecific marker for bleeding (ie, heme through a peroxidase reaction). Newer tests like FIT and sDNA have clear biological advantages over gFOBT; however, the evidence for these tests is less mature. For FIT and a select few sDNA panels, there are cross-sectional, population-based studies in which colonoscopy is the reference standard. These studies provide a reasonable estimate of one-time test characteristics, providing data for simulation modeling of the cumulative benefits, harms, and costs. Studies of stool-based micro-RNA (mi-RNA), protein and metabolic markers are nearly all case-control studies, which are considered to be preliminary or hypothesis-generating because of their tendency toward bias and need for subsequent testing using a more robust study design.
**Fecal Immunochemical Testing**

FIT uses labeled antibody to detect the globin portion of human hemoglobin. Mechanistically, this translates to a number of significant advantages over gFOBT (Table 2). Approved FIT is marketed as either qualitative or quantitative and use slightly different analytical techniques. Qualitative tests use lateral flow immunochromatographic analysis and signal positive at a preset hemoglobin cutoff. Quantitative tests directly measure hemoglobin concentration using immunoturbidimetric methods. To date, quantitative FIT has been used qualitatively with the test simply reported as “positive” above a certain threshold based on manufacturer recommendation. However, opportunity exists to leverage the quantitative data contained in the measurement.

Many studies have assessed the performance of one-time FIT using a colonoscopy gold standard to determine test characteristics for cancer and advanced adenoma in average risk screening populations. Interpretation of this literature is complicated because test performance can vary based upon the brand of the test, hemoglobin threshold defining a positive test, and number of tests applied in a single round of testing. A recently published meta-analysis summarized best available data and determined that overall sensitivity of one time FIT for cancer was 79% with a specificity of 94%.21 Among the subgroup of studies in which colonoscopy was the reference standard, cancer sensitivity was 71%. Not surprisingly, FIT sensitivity for advanced adenoma is significantly less and varies markedly depending upon the brand of test and cut-offs used. In one study comparing 6 qualitative FIT results, the 2 best performing tests had a sensitivity of approximately 25% for advanced adenoma.22
There is growing evidence regarding performance of FIT after multiple rounds of testing. In an Italian population based screening program (n = 2959), nearly 50% attended all 4 rounds of screening. Although the cancer detection rate declined over the 4 screening rounds (suggesting identification and removal of cancers over time), advanced adenoma detection remained quite consistent (approximately 33%). The results suggest that long-term adherence with FIT may subsequently translate into reduced cancer incidence through continued removal of advanced neoplasia over time.

Although our understanding of FIT’s ability to detect neoplasia has grown, significant questions remain. For example, deciding whether to use qualitative or quantitative FIT as well as the frequency and cut offs to use to define a positive test are all areas of active investigation. In this brief review it is not possible to review the evidence surrounding each of these questions, but some general statements can be made. Although there is some evidence that test characteristics comparing qualitative and quantitative FIT are quite similar, concerns regarding quality control with qualitative tests have been raised. Also quantitative testing allows the opportunity to select a test with a hemoglobin cut off that best matches available endoscopic resources. Given these types of advantages, there is growing enthusiasm for the adoption of quantitative FIT over qualitative tests.

Quantitative FIT affords the opportunity to set the test to a desired hemoglobin threshold that maximizes detection and minimizes false positive rates. That being said, determining the optimal hemoglobin threshold for screening is challenging. Factors such as age and gender influence fecal hemoglobin concentration, partly because the prevalence of neoplasia varies by these two factors. Also, studies of FIT have lacked consistency when reporting hemoglobin quantification with a given device. Efforts to standardize reporting are underway and should
improve the quality of the literature determining optimal threshold for a positive test. In the interim two recent systematic reviews suggest that lower thresholds (<20 μg of hemoglobin/g of stool) effectively balance test characteristics and improve cost effectiveness. Moving forward, it may well be possible to leverage knowledge of quantitative results to prioritize those most needing colonoscopy. A recent study identified a hemoglobin threshold of >177 μg hemoglobin/g of stool as an independent risk factor for advanced neoplasia after adjusting for factors such as sex and age (OR, 3.8; 95% confidence limit [CL], 3.07, 4.71).

Aside from research addressing the practical questions just discussed, other work to enhance understanding of FIT remains to be done. Large randomized trials directly assessing FIT versus colonoscopy as a primary tool for cancer screening are underway, but results will not be available for a decade or more. Although 1 recent study directly comparing 2 brands of FIT on important outcomes (eg, CRC mortality) has recently been reported, more comparative studies of other FIT brands and other noninvasive tests (like sDNA) are needed. The role of FIT as a screening tool in higher risk populations (eg, family history) or as an adjunct to surveillance have both recently been examined, but more work is needed to understand the utility of FIT outside of average risk screening. Finally, studies aimed at understanding how to optimize FIT-based screening to maximize adherence both with completion of the initial screen and follow-up colonoscopy when FIT positive are needed.

**Stool-Based DNA Testing**

The biological rationale is strong for measuring mutated DNA in stool. Colonocytes are continuously shed into the lumen; neoplastic cells exfoliate at a higher rate and include DNA that remains intact (ie, long) and is easier to detect than is fragmented DNA. Point mutations in oncogenes or tumor suppressor genes are specific for cancer and precancerous lesions.
Many of the published reports consists of comparing either single DNA mutations or panels of mutations in cases with CRC and controls without CRC, and may or may not include a group of subjects with precancerous (adenomatous) polyps. This literature is heterogeneous with respect to sample size, case selection and stage, DNA markers, control groups, and study findings. In contrast to evaluation of other nonhemoglobin stool-based markers, sDNA has been evaluated in population-based studies in the screening setting with colonoscopy as the reference standard. One of two such studies evaluated an early panel of DNA markers, specifically 21 point mutations involving KRAS and APC, BAT-26 (a marker of microsatellite instability), and a DNA integrity assay in 2507 asymptomatic persons 50 years of age or older (and compared the panel to Hemoccult II (Beckman Coulter, Inc, Brea, CA), a gFOBT). Cancer sensitivity was 52% for the panel vs 13% for Hemoccult II, and advanced adenoma sensitivity was 15.5% (vs 10.6% for Hemoccult II). In a parallel study of 4,482 average-risk adults, cancer sensitivity and specificity using the same marker panel were 25% and 96%, respectively. A second DNA panel that included APC, KRAS, and methylated vimentin, was tested during the latter half of this study and showed higher sensitivity (58%) but lower specificity (84%). Sensitivity for adenomas 1 cm or larger was 46%. The results from these two studies were disappointing, particularly because of the high cost of the test, which precluded repeating it at short intervals. Although this first generation sDNA test was recommended by some guideline organizations, and its use in clinical practice was undetectable.

Subsequent studies included better markers, buffers for stool stabilization, and more analytically sensitive measurement. Ahlquist and colleagues used quantitative allele-specific real-time target and signal amplification to test 4 methylated genes (NDRG4, BMP3, vimentin, and TFPI2), KRAS, and hemoglobin by immunoassay on archived specimens from 252 subjects.
with CRC, 293 colonoscopy-negative controls, and 133 subjects with adenomas 1 cm or larger. At respective specificities for CRC and large adenomas of 90% and 89%, sensitivities were 85% for CRC and 54% for large adenomas, with no difference in sensitivity between proximal and distal CRCs or large adenomas. These encouraging findings led to the pivotal study for this marker panel, in which nearly 10,000 asymptomatic persons 50 to 84 years of age were tested with commercially available FIT and a multitarget sDNA panel, which consists of molecular assays for aberrantly methylated BMP3 and NDRG4 promoter regions, mutant KRAS and β-actin (a reference gene for human DNA quantity) as well as an immunochemical assay for human hemoglobin, prior to undergoing screening colonoscopy. Quantitative measurements of each marker were incorporated into a validated logistic regression equation, with a value of 183 or more indicative of a positive test. CRC sensitivity was 92.3% for the panel and 73.8% for FIT, with respective specificities of 86.6% and 94.9%. Sensitivity for advanced, precancerous polyps was 42.4% for the panel and 23.8 for FIT. The sDNA panel was superior to FIT for detection of lesions with high-grade dysplasia (69.2% vs 46.2%, respectively) and for sessile serrated polyps 1 cm or greater (42.4% vs 5.1%, respectively). The numbers of persons who need to be screened to detect one CRC were 154 for colonoscopy, 166 for the sDNA panel, and 208 with FIT. This stool-based, multitarget panel was approved with a 3-year testing interval by the Food and Drug Administration in August of 2014, and approved for reimbursement by the Centers of Medicare and Medicaid Services shortly afterwards at a cost of $492.72 (the list price is $599). The test must be ordered from a physician’s office through direct communication with the company providing the test (Cologuard, Exact Sciences Corporation, Madison, WI).

Protein
Although gFOBT and FIT rely on detection of the protein hemoglobin as a marker of colorectal neoplasia, a number of other protein-based markers have been examined. These candidate biomarkers fall into two broad classes. Some tests rely on the detection of proteins released from products in the serum and might be expressed from inflamed and/or bleeding tissue. Other assays focus on detecting cancer-specific fecal proteins.

Calprotectin, a calcium binding protein found largely in neutrophils, has been the most widely studied of the noncancer specific proteins. However, most of the studies are small case-control studies that were not performed within the target population of interest (ie, average risk screening). One of the best evaluations of the marker was performed using stool samples obtained on individuals participating in the Norwegian Colorectal Cancer Prevention Screening trial (n = 2321) where the performance of calprotectin was compared with FIT. Calprotectin detected fewer cancers than FIT, and had lower overall specificity. Calgranulin B, one of the components of calprotectin, has separately been studied; its test characteristics do not support its use for screening.

Of the cancer-related fecal proteins, tumor M2 pyruvate kinase (M2-PK) has received the most attention as a potential stool biomarker for cancer screening. A recently reported meta-analysis summarized 10 observational studies (6 case-control and 4 cohort) of M2-PK and only included those studies where colonoscopy was performed in all study participants. The pooled CRC sensitivity and specificity were 79% and 81%, respectively. The authors specifically looked at 4 studies that compared M2-PK to FIT and found that the diagnostic odds ratio favored FIT (67.2 for FIT vs 9.5 for M2-PK, respectively).

Although there is no single fecal protein biomarker superior to FIT, there may be a role for use of these tests either in combination with FIT or applied as a panel of tests. Leen and
colleagues included a M2-PK test kit along with a FIT as part of a FIT-based program. During that screening round 13 advanced adenomas were detected and 2 were specifically found because M2-PK was added (FIT detected 11 of 13 lesions). However, the positivity rate of the M2-PK was extremely high (27%) relative to FIT (6%); the lower specificity of M2-PK would markedly increase colonoscopy use for a relatively small gain in detection.

Looking for a combination of fecal protein markers may eventually be possible. There have been improvements in technology to examine the metaproteomics of the gut through shotgun mass spectrometry of stool. Using this approach, Verberkmoes demonstrated that about 1/3 of proteins isolated in stool come from the human host and may theoretically be a biomarker target. Karl and colleagues examined 6 protein markers in 551 samples collected in a cohort that was enriched with patients with known cancer. Using Bayes logistic regression modeling, they identified a pair of markers (S100A12 and hemoglobin-haptoglobin) that demonstrated high sensitivity (88%) and specificity (95%) for cancer. Importantly, these results were not derived in a screening cohort, nor were they externally validated. Similarly, a study in 20 cancer cases and an equal number of healthy controls examined the use of fecal protein biochips with simultaneous assessment of seven different protein markers and demonstrated sensitivity for cancer of 70% and a specificity of 40%. Although these 2 studies provide some proof of principle regarding combination panel testing, such strategies are far from clinical implementation.

**RNA Testing**

Two families of RNA have been studied as markers of colorectal neoplasia. Messenger RNA (mRNAs) were investigated initially, but found to be less stable in stool than DNA. During the last several years, investigation has turned to micro-RNAs (mi-RNAs), which are 18- to 25-
nucleotide noncoding RNA molecules that regulate gene expression and translation, indirectly affecting cell differentiation, cell cycle progression, and apoptosis. More stable in stool and plasma than other nucleic acid molecules, mi-RNAs play either an oncogenic or tumor-suppressor role in the multistep process of carcinogenesis, and are thought to be cell type- and disease-specific. In a clinical setting, mi-RNA may be quantified in stool by quantitative real-time PCR (qPCR), which uses a reference gene or noncoding RNA transcript to quantify the relative amounts mi-RNA.

Although a number of studies of single mi-RNAs and panels of mi-RNAs have been published, the field is still nascent. Several technical and analytical issues affecting measurement stability remain to be addressed. The more recent use of digital PCR likely represents a technical improvement that could accelerate scientific advancement. Similar to the sDNA literature, studies on mi-RNA are largely case-control studies of subjects with CRC compared with either “healthy” or colonoscopy result-negative subjects, with or without a third group with adenomas, and is quite heterogeneous in the specific single or panel of mi-RNA, study sample size, collection methods, and preparation, making comparisons challenging. Several of these pilot/hypothesis-generating studies show discrimination, but to a modest degree. Based on systematic review and independent validation, a tissue-based, differentially-expressed mi-RNA panel has been identified from analysis of CRC tissue and paired neighboring noncancerous colorectal tissue; however, the panel has not yet been evaluated on stool specimens. Ahmed and colleagues have found differential expression of several mi-RNAs when comparing stool of cases with cancers to that of controls with no neoplasia; however, the results require independent validation. Of interest, only one differentially regulated mi-RNA (mi-106a) is
common to these two works, suggesting the need for independent validation, followed by large-scale, population-based evaluation.

**Cost Effectiveness of Noninvasive Testing**

Several cost-effectiveness analyses show that CRC screening with either invasive or noninvasive tests is cost effective, and annual FIT is often the first or second most cost-effective strategy in these models. In the cost-effectiveness analysis by Heitman and colleagues, annual FIT tests of medium and high test characteristics were the most effective and cost-effective strategies, whereas gFOBT and two (older) sDNA panels were least cost effective. More recently, Lansdorp-Vogelaar and colleagues used two microsimulation models to quantify the cost effectiveness of sDNA at a cost of $350 and used every 3 or 5 years among Medicare enrollees. Although sDNA at either interval provided fewer life-years and higher cost than other recommended CRC screening strategies, it would be cost effective at a cost of $40 to $60 at a 3-year interval, and at a cost of $350 if relative adherence to testing were at least 50% higher than other strategies.

**Conclusions**

Noninvasive testing for CRC has made great progress during the last decade and its trajectory, based on technical and analytical improvements and on active exploration of candidate markers, suggests continued rapid advancement. Both annual FIT and sDNA every 3 years are viable noninvasive strategies for average-risk CRC screening and are both endorsed by current sub-speciality guidelines. Although a large randomized trial comparing annual FIT to sDNA every 3 years programmatically would be welcomed, such a study may be impractical as both technologies are evolving rapidly and the results of any long-term study
could be outdated before completion. Optimizing use of these tests in clinical practice requires a clearer understanding of risk for advanced colorectal neoplasia among “average-risk” persons, as well as a consideration of their test characteristics for CRC and advanced, precancerous lesions, testing interval, cost, and patient preference. If progress in noninvasive marker discovery continues at its current pace, we expect further improvements in marker discrimination, higher test throughput at lower cost, and more user-friendly testing, with a resulting expected decrease in the use of colonoscopy as a primary screening test among average-risk persons. High-quality noninvasive testing is the likely conduit to achieving higher adherence to CRC screening, with anticipated continued reductions in CRC incidence and mortality.

Acknowledgments

The contents of this work do not represent the views of the Department of Veterans Affairs or the United States Government.
References


10. Lee JK, Levin TR, Corley DA. The road ahead: what if gastroenterologists were


38. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the


Table 1. Overview of Status and Supporting Evidence for Stool-Based Screening Tests

<table>
<thead>
<tr>
<th>Stool-Based Screening Test</th>
<th>Status as a CRC Screening Test</th>
<th>Examples</th>
<th>Supporting Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIT</td>
<td>FDA approved</td>
<td>Quantitative Oc Auto Micro Qualitative InSure ImmoCARE MonoHaem OC Light FIT CHEK</td>
<td>Large Scale prospective studies in screening populations</td>
</tr>
<tr>
<td>sDNA</td>
<td>FDA approved</td>
<td>Cologuard (key markers: methylation of BMP3 and NDRG4; KRAS; β-actin; hemoglobin)</td>
<td>Large scale cross-sectional studies in screening populations</td>
</tr>
<tr>
<td>Protein</td>
<td>In development</td>
<td>Non-cancer-specific Calprotectin calgranulin B, lactoferrin Cancer-specific tumorM2-pyruvate kinase Clusterin</td>
<td>Variable by marker; largely small cross sectional studies; often in populations with CRC or enriched with CRC</td>
</tr>
<tr>
<td>RNA/miRNA</td>
<td>In development</td>
<td>Up-regulated (selected) miR-21 miR-92a miR-106a miR-135b Down-regulated (selected) miR-9 miR-29b miR-143</td>
<td>Generally small scale proof of concept type studies (eg&lt;comma&gt; measuring markers in cancer tissue or stool)</td>
</tr>
</tbody>
</table>

CRC, colorectal cancer; FDA, US Food and Drug Administration; FIT, fecal immunochemical testing; mi-RNA, micro-RNA.
Table 2. Advantages of FIT Compared to Conventional Guaiac-based Fecal Occult Blood Test

<table>
<thead>
<tr>
<th>Biological Advantage</th>
<th>Clinical Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directly measures human hemoglobin</td>
<td>• Highly sensitive for cancer even with single sample testing with some devices (improved compliance)</td>
</tr>
<tr>
<td></td>
<td>• Maintains specificity even at higher levels of sensitivity (fewer false positive tests requiring definitive colon evaluation)</td>
</tr>
<tr>
<td></td>
<td>• No need to adjust diet (improved compliance)</td>
</tr>
<tr>
<td>Hemoglobin released from upper GI tract degraded in transit</td>
<td>• No need to adjust drug intake like NSAIDs or anticoagulants (improved compliance)</td>
</tr>
<tr>
<td></td>
<td>• Fewer false positives from the upper GI tract (ie, improved specificity)</td>
</tr>
<tr>
<td>Hemoglobin measurement can be quantified</td>
<td>• Definition of a positive test can be matched to colonoscopy resources</td>
</tr>
<tr>
<td></td>
<td>• Opportunity to use quantitative value to stratify risk</td>
</tr>
<tr>
<td>Hemoglobin measurement can be automated</td>
<td>• Reduces the likelihood that results are impacted by quality control issues</td>
</tr>
<tr>
<td></td>
<td>• Facilitates high throughput (eg, population based) screening efforts</td>
</tr>
</tbody>
</table>

FOBT, fecal occult blood test; GI, gastrointestinal (tract); NSAID, nonsteroidal anti-inflammatory drug.