



The importance of comparing quantitative faecal immunochemical tests (FIT) before selecting one for a population-based colorectal cancer screening programme

James E. Allison¹, Callum G. Fraser²

¹Division of Gastroenterology, University of California San Francisco School of Medicine, SFGH Campus, San Francisco, CA, USA; ²Centre for Research into Cancer Prevention and Screening, University of Dundee, Ninewells Hospital and Medical School, Dundee, Scotland, UK

Correspondence to: Dr. James E. Allison. Division of Gastroenterology, University of California San Francisco School of Medicine, SFGH Campus, 1001 Potrero Avenue, NH-3D San Francisco, CA 94110, USA. Email: james.allison@ucsf.edu.

Comment on: Gies A, Cuk K, Schrotz-King P, *et al.* Direct Comparison of Diagnostic Performance of 9 Quantitative Fecal Immunochemical Tests for Colorectal Cancer Screening. *Gastroenterology* 2018;154:93-104.

Received: 30 December 2017; Accepted: 12 January 2018; Published: 29 January 2018.

doi: 10.21037/jlpm.2018.01.05

View this article at: <http://dx.doi.org/10.21037/jlpm.2018.01.05>

Introduction: faecal immunochemical tests (FIT) for haemoglobin

FIT for haemoglobin are now commonly used in many countries as the best, currently available, non-invasive test for colorectal cancer (CRC) in asymptomatic population screening programmes (1). FIT are available in two formats, qualitative, usually based on lateral-flow immunochromatographic test strips or cassettes, and quantitative, most often based on immunoturbidimetry, and performed on small benchtop analyzers. Quantitative FIT have a number of significant advantages over qualitative FIT, a major one being that analyses of faecal samples give estimates of the faecal haemoglobin concentrations (f-Hb) (2).

Many FIT are available worldwide and, in the United States (U.S.), as of July 11, 2017, the Federal Drug Administration (FDA) Clinical Laboratory Improvement Amendments of 1988 (CLIA) test categorization database included 134 test systems for occult blood in faeces. Five are automated FIT (although results are reported only as qualitative tests) and 129 are waived non-automated FIT (3). Unfortunately, large scale comparative studies of different FIT for detection of advanced colorectal neoplasms (AN) i.e., large polyps, or any polyp with dysplasia are uncommon (4) and, the published comparisons of qualitative FIT do give cause for concern. In one such study, overall sensitivity and specificity of six FIT varied from 66%

and 96% to 92% and 62%, respectively (4). It has been documented that, although about two-thirds of the FIT used commonly in the U.S. performed acceptably on samples spiked with human haemoglobin, the low sensitivity and specificity of some meant that they probably should not be used for population-based or other screening initiatives (5).

Comparison of quantitative FIT

There are a number of comparisons of quantitative FIT and the advantages and disadvantages of approaches that can be applied in such studies have been examined in detail (6). Most published comparisons of qualitative FIT have involved assessment of only two analytical systems (7). A recent comparison of two of the three most commonly used automated FIT systems demonstrated that the OC-Sensor (Eiken Chemical Co., Ltd, Tokyo, Japan) and FOB-Gold (Sentinel Diagnostics, Milan, Italy) were equally acceptable to a screening population although FOB-Gold was more prone to have specimens submitted that were unsuitable for analysis (8). Some differences were seen. The positivity rates were different, 7.9% and 6.5% for OC-Sensor and FOB-Gold, respectively. Interestingly, the diagnostic yield of AN and positive predictive value (PPV) were not significantly different when the FIT were assessed at the same positivity rate instead of the same f-Hb cut-off. An analogous study suggested that the acceptability and diagnostic performance of HM-JACKarc (Kyowa-Medex Co., Ltd, Tokyo, Japan)

and of OC-Sensor (Eiken) systems were similar in a screening setting, but documented sound reasons for comparing FIT systems at the same positivity rate rather than at the same f-Hb cut-off (7). Such comparisons do not further investigate participants who have f-Hb less than the cut-off applied and so Gies *et al.* (9) are correct in stating that “it is unclear to what extent differences in reported sensitivities and specificities reflect true heterogeneity in test performance or differences in study populations or varying pre-analytical conditions”.

Direct comparison of diagnostic performance of nine quantitative FIT

Gies *et al.* (9) addressed these issues through a direct comparison of the sensitivity and specificity with which nine quantitative (five laboratory-based and four point-of-care) FIT detected (AN) in a single CRC screening study. This is the first study to undertake such a comprehensive simultaneous comparison of quantitative FIT. It should be noted, however, that many of the investigated FIT are not widely available and that one of the FIT widely used in Europe and Asia (HM-JACKarc) was not included. In the U.S., the FDA has not cleared, approved, or evaluated even one quantitative FIT and does not as yet accept results from large studies in other countries as proof for their use in the U.S. This means that quantitative FIT cleared by the FDA can only be used as qualitative FIT. Importantly, many of the FIT examined in the study use technology that does not facilitate rapid analysis of large numbers of faecal samples, rendering these unsuitable for large scale population-based CRC screening.

What samples should be used for FIT comparisons?

This study used faecal samples obtained from participants in Germany who were enrolled in a colonoscopy-based screening programme from 2005 through 2010 in which samples were frozen at -80°C until analysis. The faecal samples were not taken (9), as in real CRC screening application, directly from fresh faeces into the FIT specimen collection device from un-homogenized faeces, even though some of these do not enhance longer-term f-Hb stability. The faecal samples were thawed, homogenized, taken into the appropriate devices and mixed on a vortexer and then analysed. None of this is done in routine practice and this approach is not what is done in CRC screening

programmes using FIT. Nevertheless; we agree that this seems the only way to undertake a comparative study of this magnitude and, this group found only small differences when the diagnostic performance based on frozen faecal samples or faecal samples collected according to the manufacturer's instructions were compared (10).

Documentation of FIT analytical performance characteristics and units

Although the bases of the analytical methods are given in Table 1 of Gies *et al.* (9), this study gives very little information on the analytical performance characteristics of the nine FIT. These are important factors to consider when selecting a FIT for use in screening. Test calibrators and controls were said to be performed on a regular basis according to the manufacturers' instructions, but no numerical data are documented. Quantitative data on analytical performance, as documented in the FITTER standards advocated by the Expert Working Group on FIT for Screening (EWG), Colorectal Cancer Screening Committee, World Endoscopy Organization (11), should have been followed in this work.

Diagnostic performance for AN were compared at the manufacturers' pre-set f-Hb cut-offs (range, 2–17 μg Hb/g faeces), at a single uniform threshold (15 μg Hb/g faeces), and at thresholds adjusted to yield defined levels of specificity (99%, 97%, and 93%). It is well known that the sensitivity and specificity vary at different f-Hb (12) and, thus, given the different f-Hb cut-offs shown, it is hardly surprising that different clinical outcomes are obtained using the different FIT as recommended by the manufacturers.

The quoted analytical ranges for the nine FIT are very different. Could this be, in part, due to the lack of understanding as to the metrological requirements for determining and documenting these variables (13) and the lack of information provided as to whether the lower f-Hb quoted is the limit of detection (LoD) or the limit of quantitation (LoQ), as discussed in detail in a recent review in this journal (14)? Given the analytical imprecision of quantitative FIT, we believe that f-Hb should be reported as integers only and not to many significant figures after the decimal point as documented in Table 1 (9). We understand that manufacturers still quote data as ng Hb/mL buffer and guess that the authors have probably recalculated to the EWG recommended units of μg Hb/g faeces (15) from the quoted mass of faeces collected and the volume of buffer in

the specimen collection device. We strongly believe that all manufacturers, suppliers and users of FIT should use units of $\mu\text{g Hb/g faeces}$ to aid universal comprehension.

Clinical outcomes

The results given in the study of Gies *et al.* (9) are of much interest. Of the 1,667 participants who fulfilled the inclusion criteria, all 216 cases with AN and 300 randomly selected individuals without AN were included in the analysis. Not surprisingly, the sensitivities and specificities for AN varied widely when the pre-set f-Hb cut-offs or the uniform f-Hb cut-off were used. Adjusting the f-Hb cut-offs to give specificities of 99%, 97%, or 93% resulted in almost equal sensitivities for detection of AN. The authors state that this strategy gave almost equal positivity rates (2.8–3.4%, 5.8–6.1% and 10.1–10.9%, respectively).

We believe that the subjective statement of “almost equal” is misleading. Many CRC screening programmes are carried out in countries with very scarce colonoscopy resources: for 100,000 participants, 2.8% positivity would mean that 280 colonoscopies would be required whereas 3.4% would mean 340, an over 20% increase in requirement and, possibly unobtainable. The authors are correct that, in practice, determining specificity, or defining a f-Hb cut-off according to specificity, in the context of an established FIT-based screening programme is often difficult because only test-positive participants typically would be referred for colonoscopy. Although the authors state that adjusting the f-Hb cut-off to defined positivity rates would have resulted in very narrow ranges of specificities across tests, these data were not shown. It would have been of value to document these data at identical positivity rates, as recommended in recent comparisons of FIT systems (7,8). As the authors state, selecting a f-Hb cut-off to give a pre-defined positivity rate is relatively straightforward.

Discussion/conclusions

We applaud the authors for their study and agree with their statement that it is the first comprehensive comparative evaluation of diagnostic performance of a large number of quantitative FIT in a screening setting. Their Discussion section is excellent and their honesty regarding the study's weaknesses is admirable. We hope that this work stimulates other studies on the characteristics of FIT systems. As more come to market, evidence must be generated for each that, in large average risk populations, the FIT

provides acceptable performance. The work confirms that large differences in diagnostic performance variables are seen when using the f-Hb cut-offs recommended by the manufacturers and, importantly, a single uniform f-Hb cut-off does not resolve the problem. They demonstrated that the large differences that were found nearly disappeared when f-Hb cut-offs were adjusted in such a way that all the FIT achieved defined specificities at which sensitivities were also all very close. This latter point confirms the current concepts that comparison of outcomes obtained by FIT are best and easily done using identical positivity rates rather than identical f-Hb cut-offs (6). The reasons for the difference in outcomes when using one f-Hb cut-off across all systems is probably mostly due to the fact that the antibodies have different specificities and react not only to intact haemoglobin, but also to different spectra of haemoglobin degradation products.

In conclusion, studies comparing FIT performance in large average risk populations are necessary and needed. As the authors state, their results underline the need for enhanced efforts for harmonization, standardization and quality assessment of FIT. Rather than simply using the f-Hb cut-offs recommended by FIT manufacturers, screening programme organizers should choose f-Hb based on intended levels of specificity and/or manageable positivity rates consistent with the programme's colonoscopy resources.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned and reviewed by the Section Editor Dr. Li-Yi Hu (Department of Clinical Laboratory, Yongchuan Hospital of Chongqing Medical University, Chongqing, China).

Conflicts of Interest: JE Allison has no conflicts of interest to declare. CG Fraser has undertaken paid consultancy with Immunostics Inc., Ocean, NJ, USA, and Kyowa-Medex Co. Ltd, Tokyo, Japan, and has received support for attendance at conferences from Alpha Labs Ltd, Eastleigh, Hants, UK.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are

appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Young GP, Symonds EL, Allison JE, et al. Advances in Fecal Occult Blood Tests: the FIT revolution. *Dig Dis Sci* 2015;60:609-22.
2. Fraser CG, Allison JE, Young GP, et al. Quantitation of hemoglobin improves fecal immunochemical tests for noninvasive screening. *Clin Gastroenterol Hepatol* 2013;11:839-40.
3. Clinical Laboratory Improvement Amendments of 1988 (CLIA); Fecal Occult Blood (FOB) Testing. Federal Register 2017;82, No. 202.488770-3. Available online: <https://www.gpo.gov/fdsys/pkg/FR-2017-10-20/pdf/2017-22813.pdf>
4. Tao S, Seiler CM, Ronellenfitsch U, et al. Comparative evaluation of nine faecal immunochemical tests for the detection of colorectal cancer. *Acta Oncol* 2013;52:1667-75.
5. Daly JM, Bay CP, Levy BT. Evaluation of fecal immunochemical tests for colorectal cancer screening. *J Prim Care Community Health* 2013;4:245-50.
6. Fraser CG. Comparison of quantitative faecal immunochemical tests for haemoglobin (FIT) for asymptomatic population screening. *Transl Cancer Res* 2016;5:S916-S919.
7. Passamonti B, Malaspina M, Fraser CG, et al. A comparative effectiveness trial of two faecal immunochemical tests for haemoglobin (FIT). Assessment of test performance and adherence in a single round of a population-based screening programme for colorectal cancer. *Gut* 2016. [Epub ahead of print].
8. Grobbee EJ, van der Vlugt M, van Vuuren AJ, et al. A randomised comparison of two faecal immunochemical tests in population-based colorectal cancer screening. *Gut* 2017;66:1975-82.
9. Gies A, Cuk K, Schrotz-King P, et al. Direct Comparison of Diagnostic Performance of 9 Quantitative Fecal Immunochemical Tests for Colorectal Cancer Screening. *Gastroenterology* 2018;154:93-104.
10. Chen H, Werner S, Brenner H. Fresh vs Frozen Samples and Ambient Temperature Have Little Effect on Detection of Colorectal Cancer or Adenomas by a Fecal Immunochemical Test in a Colorectal Cancer Screening Cohort in Germany. *Clin Gastroenterol Hepatol* 2017;15:1547-1556.e5.
11. Fraser CG, Allison JE, Young GP, et al. Improving the reporting of evaluations of faecal immunochemical tests for haemoglobin: the FITTER standard and checklist. *Eur J Cancer Prev* 2015;24:24-6.
12. Brenner H, Werner S. Selecting a Cut-off for Colorectal Cancer Screening With a Fecal Immunochemical Test. *Clin Transl Gastroenterol* 2017;8:e111.
13. Clinical and Laboratory Standards Institute. Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures, 2nd Edition, Approved Guideline. Wayne, PA, USA: CLSI; CLSI document EP17-A2. 2012.
14. Fraser CG. Interpretation of faecal haemoglobin concentration data in colorectal cancer screening and in assessment of symptomatic patients. *J Lab Precis Med* 2017;2:96.
15. Fraser CG, Allison JE, Halloran SP, et al. A proposal to standardize reporting units for fecal immunochemical tests for hemoglobin. *J Natl Cancer Inst* 2012;104:810-4.

doi: 10.21037/jlpm.2018.01.05

Cite this article as: Allison JE, Fraser CG. The importance of comparing quantitative faecal immunochemical tests (FIT) before selecting one for a population-based colorectal cancer screening programme. *J Lab Precis Med* 2018;3:7.