Detection capability of quantitative faecal immunochemical tests for haemoglobin (FIT) and reporting of low faecal haemoglobin concentrations

Fraser, Callum G.; Benton, Sally C.

Published in:
Clinical Chemistry and Laboratory Medicine

DOI:
10.1515/cclm-2018-0464

Publication date:
2018

Document Version
Peer reviewed version

Opinion Paper

Detection capability of quantitative faecal immunochemical tests for haemoglobin (FIT) and reporting of low faecal haemoglobin concentrations

Callum G Fraser[1] and *Sally C Benton*[2]

[1] Centre for Research into Cancer Prevention and Screening, University of Dundee, Scotland, UK.

[2] NHS Bowel Cancer Screening Programme, Southern Hub, Royal Surrey County Hospital NHS Foundation Trust, Guildford, Surrey, UK.

Corresponding author: Professor Callum G. Fraser, Centre for Research into Cancer Prevention and Screening, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland, UK.

E-mail: callum.fraser@nhs.net: c.g.fraser@dundee.ac.uk

Short title: Detection capability of faecal haemoglobin

Number of words – abstract: 231 **text:** 2609

Number of Tables: none

Number of Figures: none

Number of references: 31

Keywords: detection capability; faecal immunochemical test; faecal haemoglobin; limit of blank; limit of detection; limit of quantitation

Abstract

Faecal immunochemical tests for haemoglobin (FIT) are widely used in asymptomatic population screening for colorectal (bowel) cancer. FIT are also used to assist with the assessment of patients presenting with lower abdominal symptoms. Quantitative FIT allow the generation of numerical estimates of faecal haemoglobin (f-Hb) concentrations. There is now great interest in “low” f-Hb concentrations in these clinical settings: in consequence, knowledge of the detection capability is very important for f-Hb concentration examinations. There are a number of current problems associated with the reporting of low f-Hb concentrations and wide misunderstanding of the metrological aspects of examinations of f-Hb at low concentrations. These would be solved if the detectability characteristics of f-Hb concentration examinations, namely, the limit of blank (LoB), limit of detection (LoD) and limit of quantitation (LoQ), were generated, validated and used in reporting systems exactly as recommended in the EP17-A2 guideline of the Clinical Laboratory Standards Institute (CLSI). LoB and LoD are statistical concepts, but the LoQ depends on definition of analytical performance specifications (APS). In this Opinion Paper proposals for interim APS are made, based on the current state of the art achieved with examinations of faecal samples. It is proposed that LoQ is determined at an examination imprecision of CV $\leq$10% using faecal samples naturally positive for Hb rather than faeces spiked with haemolysate. Detailed proposals for reporting f-Hb data at low concentrations are also made.

Introduction

Quantitative faecal immunochemical tests for haemoglobin (FIT) allow the generation of numerical estimates of the faecal haemoglobin (f-Hb) concentration in the samples provided for examination. FIT are currently used very widely as the best non-invasive investigation in asymptomatic population screening for bowel cancer [1]. FIT are also becoming increasingly used to assist with the assessment of patients presenting in primary care with lower abdominal symptoms who might have significant bowel disease [2]. There is great interest in low f-Hb concentrations in both these clinical settings. These low f-Hb concentrations approach the detection capabilities of the quantitative FIT systems currently available on the market. Moreover, the detection capabilities approach the currently used clinical f-Hb concentration decision limits. In consequence, an understanding of the detection capability is very important for f-Hb concentration examinations.

Current problems

Many recent publications give f-Hb concentrations in integers from zero upwards, as recently documented [3]; however, the detection capability of currently available FIT systems does not support this. In addition, some report numerical data on f-Hb concentrations with significant figures after the decimal point, as in the recent comparison of nine quantitative FIT approaches by Gies et al [4]. Both of these strategies for reporting f-Hb concentrations seem inappropriate, because neither the examination imprecision achievable nor the detection capabilities of currently available FIT systems warrant them. In part this might be due to the fact that some

manufacturers and users of FIT still quote their f-Hb concentration data as ng Hb/ml buffer and authors probably recalculate, from the quoted mass of faeces collected and the volume of buffer in the specimen collection device, to units of µg Hb/g faeces [5]. All manufacturers, suppliers and users of FIT should use µg Hb/g faeces to aid universal comprehension and transferability of data across FIT systems, as recommended by the Expert Working Group on FIT for Screening, Colorectal Cancer Screening Committee, World Endoscopy Organization (EWG) [6].

As discussed in detail recently [7], an additional problem is that there is wide misunderstanding of the metrological aspects of examinations of f-Hb at low concentrations. The terminology used in the literature from both manufacturers and suppliers of FIT systems is often incorrect and misleading, such as the wide use of the term “sensitivity”. Moreover, many publications on the use of FIT also use inappropriate terms and, sometimes, the numerical data documented about the FIT system used are actually reported incorrectly [3]. For example, in the recent study of Grobbee et al [8], the examination performance characteristics as documented by the manufacturers are actually misquoted. In part, this is understandable given the diversity of current terms used for the lowest f-Hb concentration that can be determined. There are a number of conflicting guidelines and recommendations on detection capability of clinical measurements, including from professional bodies, and as such, understanding may be inhibited by the fact that many of the metrological and technical terms used to describe processes for evaluating methods vary in different sectors, both in their meaning and the way they are determined [9], which does not help clarity.

A solution

Now seems to be the right time for all involved in generation and application of f-Hb concentration data, namely, FIT system manufacturers and suppliers, academic researchers, research funding bodies, authors and reviewers of papers, reviews and materials in modern media, journal editors and professionals in laboratory medicine to all use a single vocabulary and set of approaches. This opinion paper expands on the proposed approaches published recently [7]. These strongly advocated the use of internationally accepted terminology for the detection capability of examinations used in laboratory medicine and the derivation and application of the relevant performance characteristics.

Since quantitative estimates of f-Hb concentration are probably best determined in medical laboratories accredited to ISO 15189 [10], we propose that the recommendations promulgated by the Clinical and Laboratory Standards Institute (CLSI), supported by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) [11], should be applied.

Terminology

The recommended terms to describe the detection capability of FIT are limit of blank (LoB), limit of detection (LoD) and limit of quantitation (LoQ). The details of how these are correctly established by manufacturers and validated by users, if required, are documented in detail in CLSI EP17- A2, Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures, 2nd Edition, 2012 [11]. Full examples of the correct methodology to determine these performance characteristics are

comprehensively documented. In addition, an excellent simple guide is given in a document prepared by a well-known manufacturer in laboratory medicine: this is available on the Internet [12].

Limit of blank: The LoB is the highest measured result (or analytical signal) likely to be observed (typically at 95% certainty) for a sample containing no f-Hb (a blank sample). It is the highest result that could be observed when a blank sample is repeatedly analysed. LoB is determined by estimating the standard deviation (SD) of replicate analyses of sample containing no f-Hb (blanks). Because allegedly everyone has some blood in their faeces, even in very tiny concentrations, we advocate the use of replicate analysis of the buffer in the specimen collection devices of the FIT system be used to determine the LoB rather than faecal samples.

Limit of detection: LoD is the lowest concentration at which f-Hb can be detected 95% of the time. LoD is determined by first determining LoB and then performing studies involving generation of replicate analyses of a sample or samples of faeces containing a very low f-Hb concentration. Mathematically, LoD can be calculated as LoD = LoB + (1.645 × the analytical SD of samples with low f-Hb concentration): 1.645 is used for 95% probability because this is the appropriate one-sided Z-score. Here, it would be best to use real faecal samples obtained from participants in screening or patients presenting with lower abdominal symptoms, which should be collected into the FIT system specimen collection devices. Specimens collected into traditional faecal collection pots are unsuitable for the determination of LoD, since any f-Hb will have degraded [13, 14] and the faeces will contain a mixture of f-Hb and degradation products, some of which might react with the polyclonal antibodies

usually used in FIT. In consequence, it would be the LoD of this heterogeneous mixture that would be being estimated and not the LoD for f-Hb per se.

Limit of quantitation: LoQ is the lowest amount of f-Hb that can be reliably measured. Whereas LoB and LoD are determined using statistical approaches, definition of LoQ depends on the documentation of pre-defined examination acceptance criteria. Practically, LoQ is the lowest f-Hb concentration that can be determined when some predefined analytical performance specifications (APS) are satisfied. The APS should be established using an internationally accepted and well-documented strategy. There are a number of methods for determining LoQ. Use of the “precision profile” in which imprecision is plotted versus the f-Hb concentration has many advantages, allowing users to set LoQ based upon their own objectively set APS if they wish to use different APS to those recommended by the manufacturer of the FIT system used, or the APS proposed here.

Setting analytical performance specifications (APS) for LoQ

Definition of what is acceptable examination performance, through the setting of objective APS, is necessary to document the LoQ. APS are widely defined as: the examination performance characteristics that are required to facilitate optimal health care. The setting of APS has been the subject of much research over the last 40 years [15]. Recently, the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) published a consensus statement on defining APS [16].

The consensus agreed three different models to set APS and the statement details their advantages and disadvantages

.

Model 1. *Based on the effect of examination performance on clinical outcomes.*
EFLM state that this can be done, in principle, using different types of studies: direct outcome studies – investigating the impact of examination performance on clinical outcomes - and indirect outcome studies – investigating the impact of examination performance on clinical classifications or decisions.

Model 2. *Based on components of biological variation of the measurand*
The advantage of this model is that it can be applied to most measurands for which estimates of the components of biological variation are available.

Model 3. *Based on state-of-the-art*
EFLM state that this could relate to the highest level of examination performance technically achievable. Alternatively, it could be defined as the performance achieved by a certain percentage of laboratories.

Clearly, the three models use very different principles. Moreover, it is evident that some models will be better suited for certain measurands than for others. A list has been made by EFLM allocating measurands to different models [17], but f-Hb is not included. In addition, it is important to note that APS generated using these models may not be able to be met with currently available methodology and technology. Thus, APS may be aspirational rather than operational.

Assessing the use of these three models for examinations of f-Hb concentrations in turn, there are no studies on the effect of performance on clinical outcomes, although mathematical models such as those documented by Petersen [18] could be

generated using the consequences of performance on the distributions of f-Hb concentrations and, thereby, on clinical outcomes. There are no data on the biological variation of f-Hb concentrations, although estimates could be generated from the data collected on f-Hb concentrations in screening programmes that use two or more samples, in which the asymptomatic population could be regarded as apparently healthy [19], or studies done as that recently documented for faecal calprotectin [20]. However, the variation involved in sample collection, transport and handling (pre-examination variation) would likely have impact on the estimates generated. Thus, at this time, it seems that interim APS will have to be based upon the state of the art.

The state of the art of faecal haemoglobin concentration examinations

Unfortunately, very few publications actually follow the EWG guidelines on standards for FIT evaluation reporting guidelines, namely the FITTER guidelines [21,22]. Consequently, there is a paucity of data on examination performance characteristics in the peer-reviewed literature. Some data are available on examination performance attained, such as those documented in a comparative evaluation of four FIT systems [23]: these were generated using dilutions of a haemolysate of venous blood and thus do not represent those attained with real faecal samples and, in addition, the nomenclature used did not follow the CLSI recommendations. Further data are documented in a study on strategies to conduct evaluations of FIT [24], but again these were obtained using artificial biological samples (ABS), namely dilutions of a haemolysate of venous blood and reconstituted lyophilised third-party quality control materials Such quality control materials, provided by the manufacturers of FIT systems, have also been used to monitor the quality of f-Hb concentration examinations during a few of the studies on the use of FIT in the assessment of

symptomatic patients [25-27]. None of these data use replicate analysis of faecal samples provided by participants or patients and are, therefore, likely to underestimate the examination imprecision attained. However, preliminary work from the laboratory of one of the authors (SCB) has used such faecal samples and created imprecision profiles. Taking all of the published data on examination imprecision into account, we recommend an interim APS for examination reproducibility of CV $\leq$10%.

The guidelines on setting APS for LoQ suggest that these should be defined for bias. It is not considered that this can be done objectively at this time. As discussed earlier, the available FIT systems use polyclonal antibodies which react, not only with intact haemoglobin, but also early degradation products. Thus, the measurand is not identical across FIT systems. However, the Working Group on FIT of the Scientific Division of the IFCC (IFCC SD WG-FIT) is progressing the attainment of traceability of results of f-Hb concentration examinations to higher metrological materials and methods, with the aim of enhancing comparability across FIT systems [28] This will allow bias to be minimised and, in consequence, the APS for measurement uncertainty is also $\leq$10%.

Proposals for reporting f-Hb data at low concentrations

- **Proposal 1:** f-Hb concentrations should not be reported to more significant figures than whole integers.

- **Proposal 2**: f-Hb concentrations less than the LoD should be termed “not detected” or “undetectable”.

- **Proposal 3:** manufacturers should make precision profiles available to all users and detail their derivation.

- **Proposal 4:** For academic use: f-Hb concentrations greater than the LoD could advantageously be documented, but it should be ensured that the correct LoD, as recorded by the manufacturer or supplier of the FIT system, is clearly detailed in all publications. Alternatively, if the laboratory generates the LoD from evaluation studies done *in situ*, detail of these need to be documented.

- **Proposal 5:** Such academic reports should follow the FITTER guidelines [21,22] and inform on examination performance characteristics achieved, particularly at or near the LoD.

- **Proposal 6**: For routine clinical use: numerical f-Hb concentrations should be reported only when greater than the LoQ, defined by the manufacturer according to CLSI EP17-A2 [11] and validated by the laboratory if required for accreditation purposes: f-Hb concentrations less than the LoQ (x) should be reported as < x µg Hb/g faeces.

- **Proposal 7**: If a more sophisticated reporting system is required, one suggested option is [11,12]: report as

f-Hb concentration < LoD = not detected

f-Hb concentration LoD < result < LoQ = f-Hb detected

f-Hb concentration ≥ LoQ = report the found f-Hb concentration

- If LoD < result < LoQ, more sophisticated users of results might appreciate a report such as f-Hb = x µg Hb/g faeces with a comment such as “interpret this result with caution due to higher examination imprecision” or similar. Further more complex options are detailed in CLSI EP17-A2 [11].

- **Proposal 8:** Efforts should be made to communicate the correct interpretation of reports of f-Hb concentration examination results to users and efforts should be made to encourage professionals in laboratory medicine to become involved with the other health care professionals involved in all uses of FIT [29].

Further requirements

Quality management techniques should be in place to monitor examination performance, including use of internal quality control materials, incorporating third-party controls, at appropriately low f-Hb concentrations. External quality assessment schemes (EQAS) are also urgently required. An interesting dilemma for the providers of EQAS for f-Hb concentrations is to decide whether mock faecal matrices with added haemoglobin should be circulated with the users having to sample into the appropriate specimen collection devices used by them: however, such EQAS would be assessing pre-examination variation as well as examination variation. Interestingly, a very recent paper describes the development of a ready to use

artificial faeces containing Hb and glycerol as an internal standard: it was concluded that the in-house performance characteristics suggested that this artificial faeces was acceptable as an EQAS material for FIT [30]. Circulation of simple lyophilised, or liquid stable, materials containing human haemoglobin would assess only the examination variation. Data from both types of EQAS would be of interest: few exist at present but the IFCC SD WG-FIT is planning to collate and make available a list of available EQAS for f-Hb concentration examinations.

Quality management procedures should be in place to ensure consistency of performance at low f-Hb concentrations when lots of reagents are changed with pre-set criteria for acceptance or rejection of lots: these have been termed acceptance quality checks [31].

Conclusions

The detectability characteristics of faecal haemoglobin concentration examinations should be generated, validated and used in reporting systems exactly as recommended in CLSI EP17-A2 [11]. Proposals for application of the detectability characteristics of LoD and LoQ are made in this Opinion Paper, along with the necessary definition of the APS required for documentation of LoQ.

Author contributions: Both authors accept responsibility for the entire content of this submitted manuscript and have approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: CGF undertook consultancy with Kyowa-Medex Co. Ltd, Tokyo, Japan, and received remuneration from Alpha Labs Ltd to support participation in conferences. SCB: none declared.

References

1. Young GP, Symonds EL, Allison JE, Cole SR, Fraser CG, Halloran SP, et al. Advances in fecal occult blood tests: the FIT revolution. Dig Dis Sci 2015;60:609-22.

2. Westwood M, Lang S, Armstrong N, van Turenhout S, Cubiella J, Stirk L, et al. Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. BMC Med 2017;15:189.

3. Fraser CG. Faecal haemoglobin concentration and personalised assessment of the risk of colorectal neoplasia. J Lab Precis Med 2017;2:71.

4. Gies A, Cuk K, Schrotz-King P, Brenner H. Direct comparison of diagnostic performance of 9 quantitative fecal immunochemical tests for colorectal cancer screening. Gastroenterology 2018;154:93-104

5. 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.

6. Fraser CG, Allison JE, Halloran SP, Young GP. A proposal to standardize reporting units for fecal immunochemical tests for hemoglobin. J Natl Cancer Inst 2012;104:810-4.

7. 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.

8. Grobbee EJ, Schreuders EH, Hansen BE, Bruno MJ, Lansdorp-Vogelaar I, Spaander MCW, et al. Association between concentrations of hemoglobin determined by fecal immunochemical tests and long-term development of advanced colorectal neoplasia. Gastroenterology 2017;153:1251-9.e2.

9. Magnusson B, Örnemark U (eds). Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, 2nd ed., 2014.

10. ISO 15189:2012. Medical laboratories -- Requirements for quality and competence. http://www.iso.org/iso/catalogue_detail?csnumber=56115

11. 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. https://clsi.org/standards/products/methodevaluation/documents/ep17/

12. Beckman Coulter. Information bulletin. Understanding detection capability: LoB, LoD and LoQ in the clinical laboratory. Available online: https://www.beckmancoulter.com/ucm/idc/groups/validatedcustomer/@wsr/@literature/documents/document/glb_bci_155969.pdf

13. Brown LF, Fraser CG. Effect of delay in sampling on haemoglobin determined by faecal immunochemical tests. Ann Clin Biochem 2008;45:604-5.

14. Mellen S, de Ferrars M, Chapman C, Bevan S, Turvill J, Turnock D. Evaluation of sample stability for a quantitative faecal immunochemical test and comparison of two sample collection approaches. Ann Clin Biochem 2018. [Epub ahead of print].

15. Fraser CG. The 1999 Stockholm Consensus Conference on quality specifications in laboratory medicine. Clin Chem Lab Med 2015;53:837-40.

16. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med. 2015;53:833-5.

17. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med. 2017;55:189-94.

18. Petersen PH. Performance criteria based on true and false classification and clinical outcomes. Influence of analytical performance on diagnostic outcome using a single clinical component. Clin Chem Lab Med 2015;53:849-55.

19. Fraser CG, Sandberg S. Biological Variation. Chapter 5, in Tietz Textbook of Clinical Chemistry and Molecular Diagnostics, 6th Edition, Saunders, 2018.

20. Calafat M, Cabré E, Mañosa M, Lobatón T, Marín L, Domènech E. High within-day variability of fecal calprotectin levels in patients with active ulcerative colitis: what is the best timing for stool sampling? Inflamm Bowel Dis 2015;21:1072–6.

21. Fraser CG, Halloran SP, Allison JE, Young GP. Making colorectal cancer screening FITTER for purpose with quantitative faecal immunochemical tests for haemoglobin (FIT). Clin Chem Lab Med 2013;51:2065-7.

22. Fraser CG, Allison JE, Young GP, Halloran SP, Seaman HE. Improving the reporting of evaluations of faecal immunochemical tests for haemoglobin: the FITTER standard and checklist. Eur J Cancer Prev 2015;24:24-6.

23. Carroll MRR, Piggott C, Pearson S, et al. Evaluation of quantitative faecal immunochemical tests for haemoglobin, Guildford Medical Device Evaluation Centre. 2012. http://www.worldendo.org/assets/downloads/pdf/activities/fit_reports/gmec_fit_evaluation_report.pdf

24. Rubeca T, Cellai F, Confortini M, Fraser CG, Rapi S. Impact of preanalytical factors on fecal immunochemical tests: need for new strategies in comparison of methods. Int J Biol Markers 2015;30:e269-74.

25. McDonald PJ, Digby J, Innes C, et al. Low faecal haemoglobin concentration potentially rules out significant colorectal disease. Colorectal Dis 2013;15:e151-9.

26. Godber IM, Todd LM, Fraser CG, MacDonald LR, Younes HB. Use of a faecal immunochemical test for haemoglobin can aid in the investigation of patients with lower abdominal symptoms. Clin Chem Lab Med 2016;54:595-602.

27. Auge JM, Rodriguez C, Espanyol O, et al. An evaluation of the SENTiFIT 270 analyser for quantitation of faecal haemoglobin in the investigation of patients with suspected colorectal cancer. Clin Chem Lab Med 2018;56:625-33.

28. Benton SC. IFCC - FIT Working Group (FIT-WG). IFCC e-news 2017:16-7. http://www.ifcc.org/media/461890/IFCCeNewsJune2017.pdf

29. Fraser CG, Allison JE, Young GP, Halloran SP. Newer fecal tests: opportunities for professionals in laboratory medicine. Clin Chem 2012;58:963-5.

30. Yasui R, Yamada M, Takehara S, Sakurabayashi I, Watanabe K. Novel artificial stool material for external quality assurance (EQA) on a fecal immunochemical test for hemoglobin (FIT): The confirmed utility of stable hemoglobin and an internal standard material. Clin Chim Acta 2018;483:76-81.

31. McDonald PJ, Anderson CM, Fraser CG. Acceptance quality checks for qualitative fecal immunochemical tests ensure screening program consistency. Int J Cancer 2011;128:2