Detection of advanced colorectal neoplasia and relative colonoscopy workloads using quantitative faecal immunochemical tests: an observational study exploring the effects of simultaneous adjustment of both sample number and test positivity threshold

Graeme P Young [ID] ,[1] Richard J Woodman,[2] Erin Symonds[3,1]

For numbered affiliations see end of article.

Correspondence to
Professor Graeme P Young;
graeme.young@flinders.edu.au

ABSTRACT

Objective When screening for colorectal cancer (CRC) using quantitative faecal immunochemical tests (FIT), test parameters requiring consideration are the faecal haemoglobin concentration (f-Hb) positivity cut-off and the number of stools sampled. This observational study explored variation in f-Hb between samples and the relationship between sensitivity for advanced neoplasia (AN, cancer or advanced adenoma) and colonoscopy workload across a range of independently-adjusted parameter combinations.

Design Quantitative FIT data (OC-Sensor) were accessed from individuals undergoing personalised colonoscopic screening with an offer of 2-sample FIT in the intervening years. We estimated variation in f-Hb between samples in 12 710 completing 2-sample FIT, plus test positivity rates (colonoscopy workload) and sensitivity for AN according to parameter combinations in 4037 instances where FIT was done in the year preceding colonoscopy.

Results There was large within-subject variability between samples, with the ratio for the second to the first sample f-Hb ranging up to 18-fold for all cases, and up to 56-fold for AN cases. Sensitivity for AN was greatest at lower f-Hb cut-offs and/or using 2-sample FIT. Colonoscopy workload varied according to how parameters were combined. Using different cut-offs for 1-sample FIT and 2-sample FIT to return similar sensitivity, workload was less with 2-sample FIT when the sensitivity goal exceeded 35%.

Conclusion Variation in f-Hb between samples is such that both parameters are crucial determinants of sensitivity and workload; independent adjustment of each should be considered. The 2-sample FIT approach is best for detecting advanced adenomas as well as CRC provided that the colonoscopy workload is feasible.

INTRODUCTION

Many countries use faecal occult blood tests (FOBT) for colorectalcancer (CRC)

Summary box

What is already known about this subject?

► Quantitative faecal immunochemical tests allow flexibility in screening for colorectal neoplasia but because there is day-to-day variation in the concentration of haemoglobin in stools, sampling just one stool compared with two increases the risk of missing a significant lesion, especially advanced adenomas which bleed less than cancers. Most organised screening programmes around the world adjust the cut-off for test positivity so as to control the colonoscopy workload when using just one sample, but there has been little consideration of the workload implications when positivity cut-off is assessed in relation to the number of samples collected.

What are the new findings?

► Due to the very large variation in faecal haemoglobin concentration demonstrated between samples, many advanced adenomas were missed when sampling just one stool and especially when using a higher cut-off for test positivity. The findings demonstrated that a 1-sample test was not necessarily associated with a lower colonoscopy workload if a 2-sample test had its cut-off set independently to return a comparable sensitivity. If detection of advanced adenomas is an important goal of the screening programme and where a colonoscopy workload of >5% is feasible, a 2-sample test required fewer colonoscopies for an equivalent sensitivity.

screening in the form of either guaiac FOBT (gFOBT) or the newer faecal immunochemical tests for haemoglobin (FIT).[1] The development of quantitative FIT has improved detection of colorectal neoplasia and is now

Summary box

How might it impact on clinical practice in the foreseeable future?
► Screening programmes wishing to reduce colorectal cancer incidence through detection of advanced adenomas, should carefully consider workloads associated with a 2-sample test as this is likely to provide better detection of advanced adenomas with a lower or equivalent colonoscopy workload.

the predominant FOBT screening technology used worldwide.[1–3] Recent studies have shown that FIT reduces CRC mortality as well as incidence through detection and removal of early stage cancers or precancerous adenomas.[4–6]

Quantitation of faecal haemoglobin concentration (f-Hb) has provided flexibility for screening programmes to choose the criterion value (the 'cut-off' f-Hb at which a test is reported as positive) that triggers diagnostic colonoscopy.[2 7–11] This facilitates planning of health service capacity to manage the resultant colonoscopy workload, but choice of cut-off also affects sensitivity for neoplasia.[10 12–14]

The positivity cut-off is not the only variable to be considered as it is known that blood loss from neoplasms and hence f-Hb varies from day to day.[15 16] Traditionally, gFOBT-based screening has required collection of three stool samples to allow for this[16] but the better sensitivity of FIT for CRC and advanced adenomas[17 18] has resulted in a shift to sampling just one or two stools. Japan introduced 2-sample FIT screening in 1992 on the basis of a randomised controlled trial.[19] Within a few years studies in Japan had shown a reduced mortality when screening with FIT compared with gFOBT.[20] Most national organised population screening programmes now request sampling from just one sample[13] although several regions including Australia,[21] Slovenia[22] and certain Canadian provinces[3] elect to use a 2-sample approach.

The underpinning biological determinants for detection of a cancer or advanced adenoma by FIT are how much they bleed and how much it varies from day to day, while the technological determinant is the positivity cut-off chosen. Both test parameters—the cut-off and number of stools sampled—can be varied independently when using quantitative FIT[8 23] and each should be carefully considered as there will be implications for key outcomes of a screening programme.[14 24–26] Currently applied cut-offs vary from 15 to higher than $80\,\mu g$ Hb/g faeces[3] within organised screening programmes. Selecting a lower cut-off achieves better sensitivity for advanced neoplasia (cancer and advanced adenomas)[7 9] but screening jurisdictions vary in how low they are prepared to go as lowering it is associated with a concomitant increase in the positivity rate, leading to a higher colonoscopy workload.[9] Many countries have also chosen one sample as there is concern that requesting people to sample two rather than one stool might affect the participation rate

which will in turn affect intention-to-screen outcomes. Given that f-Hb varies between stools, one sample could be positive while the other negative. Such disagreement in test positivity rate means that sensitivity for neoplasia will vary between stools depending on the chosen positivity cut-off.

As a result, shifting from a 2-sample test to a 1-sample test (or vice versa) affects sensitivity for neoplasia and colonoscopy workload (positivity rate in the screened population) unless the positivity cut-off is also adjusted.[9] There has been little consideration of the approach where sample number and cut-off are simultaneously but independently adjusted. Choice of test parameters (variables of sample number and criterion value for the positivity cut-off) relate directly to the strategic goals of a programme, such as targeting a sensitivity for advanced adenomas that reduces incidence of CRC, choosing a feasible colonoscopy workload or compromising between the two with acceptable cost-effectiveness.

This study assessed sensitivity of FIT for advanced colorectal neoplasia and the associated population test positivity rates across a range of combinations of the two test parameters (each adjusted independently). We also assessed variation in f-Hb between samples and how this affected agreement (reproducibility) in test positivity at different cut-offs.

METHODS
Overview
FIT and colonoscopy data were reviewed from a large population—the 'SCOOP' programme[27–29]—at increased risk for CRC and under surveillance with intermittently-scheduled colonoscopy plus a 2-sample FIT in the intervening years. Data from those completing FIT were analysed to determine within-individual variability between the first and second FIT, as well as population positivity rates for combinations of 1-sample FIT and 2-sample FIT with different f-Hb cut-offs. Data from a subgroup who had completed FIT within 12 months prior to colonoscopy were analysed for sensitivity for advanced neoplasia modelled across a wide range of cut-offs (10 to $150\,\mu g$ Hb/g faeces) in the settings of 1-sample testing and 2-sample testing.

Population
Data were extracted for people enrolled in the SCOOP programme during the period July 2008 to April 2019. Enrolment was determined by their physician according to whether their risk (based on personal or family history of neoplasia) was sufficient to justify intermittent colonoscopy (usually third or fifth yearly) according to Australian guidelines.[30] In addition to colonoscopy, individuals were given the opportunity to complete 2-sample FIT every 1 to 2 years between scheduled colonoscopies. The indication for being in the programme was personal history of colorectal neoplasia in 68.1% (13 095/19 229) and family history of CRC in all others.

All cases completing a 2-sample FIT were included for assessment of within-individual variability of f-Hb between the two samples, as well as estimation of population test positivity rate. Cases eligible for inclusion in the estimates of test sensitivity were those who had undergone a colonoscopy (regardless of the FIT result) and who had returned a 2-sample FIT in the preceding 365 days. Cases were excluded as follows: incomplete colonoscopy, poor preparation, uncertainty about the diagnosis, diagnosis of inflammatory bowel disease or a genetic syndrome (familial adenomatous polyposis or Lynch syndrome), or failed sample development.[31]

Programme interventions

FIT kits (OC-Sensor, Eiken Chemical Company, Tokyo, Japan) were provided by mail, with instructions as previously described.[32] Participants were instructed to sample from two different stools (passed at least 30 min apart), to keep the samples in a cool place (eg, refrigerator) and to return the collected samples, with collection date and time, within 2 weeks. Determination of f-Hb followed manufacturer instructions as previously described[32] and results expressed as µg Hb/g faeces. When a FIT result was positive (cut-off 20 µg Hb/g faeces (100 ng/mL sample buffer)) in either stool, the scheduled colonoscopy was brought forward.[33]

Colonoscopies were scheduled within participating institutions (Flinders Medical Centre, Bedford Park; Repatriation General Hospital, Daw Park; Noarlunga Hospital, Noarlunga Centre; South Australia) and were conducted according to best practice and accreditation requirements at the time. Pathological findings were characterised by colonoscopic appearance and histopathology.[33]

Outcomes and modelling

FIT positivity for the entire population was estimated across a range of cut-offs up to 150 µg Hb/g, using either the 1-sample (specifically the first sample) or 2-sample (highest in either sample) approach. *Colonoscopy workload* was defined as the FIT positivity rate for the entire population who returned a FIT according to the various cut-off and sample-number combinations. The *sensitivity* for advanced neoplasia in cases completing colonoscopy within 12 months of providing a 2-sample FIT, was determined as the proportion of cases returning a positive FIT at the different cut-off/sample number combinations.

Case phenotype for a given colonoscopic examination was defined as the most advanced pathology if multiple neoplasms were present, with the hierarchy being CRC, advanced adenoma, non-advanced adenoma and cases without neoplasia. Cancer was defined by presence of invasive colorectal adenocarcinoma. Advanced adenoma was defined by presence of adenoma with any of the following features: size ≥10 mm, high grade dysplasia or villous change, ≥3 adenomas of any size or type, sessile serrated adenoma with dysplasia or traditional serrated

adenoma. Non-advanced adenoma consisted of any other adenoma state.

Statistical analysis

Agreement in f-Hb values between first and second samples was assessed using Bland-Altman analysis, with log transformation of the haemoglobin (Hb) values prior to analysis due to their positive skew. The upper and lower 95% LOA of the Bland-Altman plots provided the estimated degree to which 95% of subjects' log-transformed Hb values differed between first and second samples. The upper and lower tolerance limits which represent the 95% CIs for the upper and lower LOA were also calculated.

The frequency of agreement of a FIT result between the first and second samples was assessed using logistic regression with agreement as the dependent variable, phenotype as the independent variable and using different cut-off values for agreement.

Test positivity rate across cut-offs (agreement within phenotypes) was assessed using Poisson regression with the total number positive as the dependent variable, cut-off as the independent variable and the total number of tests as the exposure. Results are expressed as a rate ratio for positivity at each cut-off versus a cut-off value of 10 µg Hb/g.

Log-transformed f-Hb levels were compared between phenotypes using linear regression. Back-transformation of the estimated coefficients was then used to provide the ratio of the geometric mean Hb between phenotypes.

All 95% CIs for positivity and sensitivity were calculated using the Exact method for a finite population and based on size of the group and of the relevant population from which it was drawn. Analyses were performed using Stata V.16.0 (StataCorp, USA), Excel and Prism. All hypothesis tests were two-sided with a Type 1 error rate of alpha=0.05.

Role of the funding source

Eiken Chemical Company provided test kits only and had no other roles including none in the study design, conduct or manuscript preparation. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Participants

Of the 19 229 registrants in the SCOOP programme, a 2-sample FIT was returned on 50 396 occasions. FIT was completed correctly by 12 710 on 32 413 occasions, and 7.04% were positive considering a 2-sample test (either positive) and a cut-off for positivity of 20 µg Hb/g. Colonoscopies were undertaken on 18 717 occasions; 43.3% (8104/18 717) of these had a finding of cancer or adenoma.

Due to the study duration, some people underwent several colonoscopies; there were 4244 colonoscopies conducted in 3499 people within 365 days of completing
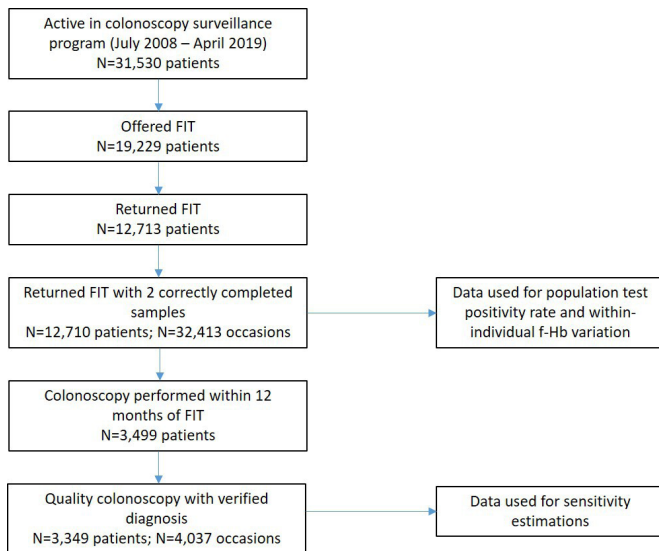
Figure 1 Details of participant engagement in surveillance and nature of the subpopulations used for modelling. FIT, faecal immunochemical test; f-Hb, faecal haemoglobin concentration.

the 2-sample FIT (figure 1). After exclusions, 4037 were eligible for outcome analysis. Advanced neoplasia was diagnosed in 15.5% (n=626) of the colonoscopies, which included 0.5% with CRC (n=21) and 15.0% with advanced adenomas (n=605). Adenomas of any type were detected in 44.7% (n=1804).

Faecal haemoglobin concentrations

Clinical phenotypes of eligible cases with their demographic characteristics and f-Hb (for each sample and for the maximum of the two), are shown in table 1. The f-Hb

in cancer cases were significantly higher than those in each other phenotypic group (p<0.001) and significantly lower in those with a normal colorectum than each other phenotype (p<0.001), regardless of which sample context was examined. The geometric mean of f-Hb in cases with advanced adenoma was 21.0% higher (Exp(β)=1.210; 95% CI=0.978 to 1.498) than those with low-risk adenoma (p=0.08, 2-sample FIT based on the higher value).

Comparison of haemoglobin concentration between first and second samples

Analysis of agreement (by Bland-Altman analysis) in f-Hb between the first and second samples is shown in online supplemental figure 1. There was no systematic fixed bias favouring either the first or second sample; mean difference in Hb between samples (all samples) was 0.003, 95% CI=-0.018 to 0.024 (log$_e$ µg Hb/g faeces), p=0.799, and for advanced neoplasia was 0.005, 95% CI -0.179 to 0.170, p=0.959. There was large within-subject variability between samples, with the ratio for the second to the first sample f-Hb ranging up to 18-fold for all cases, and even more widely for AN cases (up to 56-fold) (online supplemental table 1 and figure 1).
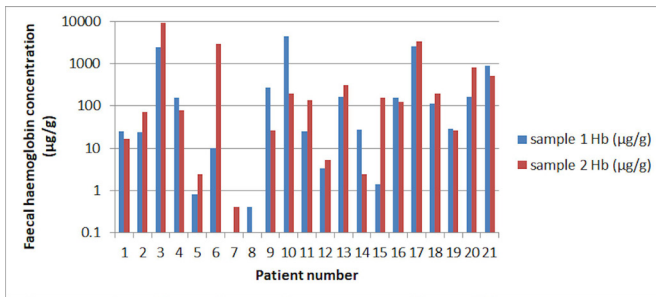
Actual f-Hb concentrations for first and second samples are shown for patients with cancer in figure 2. The proportion of cases that would be called positive with a 1-sample FIT, given a cut-off currently in use within screening programmes, would vary widely; the same would apply to advanced adenoma (data not shown).

Result concordance between samples

Due to the large variation in f-Hb between samples, we also examined the frequency of *agreement of the result*

Table 1 Clinical phenotypes of eligible cases with their demographic characteristics and faecal haemoglobin concentrations (for each sample and for the maximum of the two)

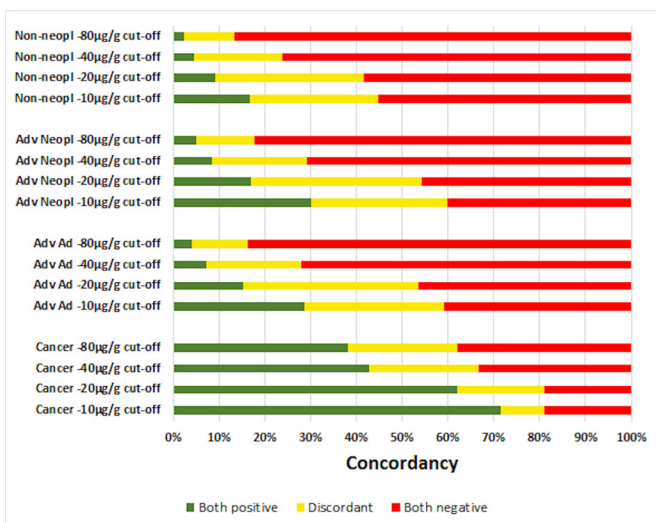|	Main phenotype				
	Advanced neoplasia (cancer plus advanced adenoma)	Cancer	Adenoma		
			Advanced	Non-advanced	No neoplasia
n=4037	626 (15.5%)	21 (0.5%)	605 (15.0%)	1199 (29.7%)	2212 (54.8%)
Demographic					
Age (years), Median, IQR	66.5 (60.9 to 70.6)	69.8 (64.0 to 71.1)	66.4 (60.7 to 70.6)	65.3 (59.1 to 70.2)	63.4 (56.0 to 68.9)
Gender (% male)	356 (56.9)	10 (47.6)	346 (57.2)	666 (55.5)	992 (44.8)
Faecal haemoglobin by sample context					
First sample (µg/g, median IQR))	6.0 (0.8 to 29.9)	29.0 (10.0 to 165)	5.6 (0.8 to 28.4)	3.2 (0.4 to 22.9)	1.6 (0 to 20.6)
Second sample (µg/g, median IQR))	6.6 (1.2 to 29.0)	126.2 (13.9 to 358.9)	5.6 (1.0 to 28.2)	3.2 (0.4 to 24.0)	1.6 (0 to 20.2)
Highest of either sample (µg/g, median IQR))	22.4 (2.0 to 49.2)	155.0 (27.4 to 837.1)	21.6 (1.8 to 44.7)	11.4 (1.2 to 42.4)	4.2 (0.6 to 38.5)

Figure 2 Faecal haemoglobin concentrations for first and second samples in patients with colorectal cancer. Hb,haemoglobin.

(concordancy) in the first and second samples at selected cut-off values in each phenotype (figure 3). Based on logistic regression, there were significant differences in concordancy across the range of cut-offs for all phenotypes (p<0.001 for each). At each cut-off, a positive result for both samples was more likely in cases with cancer compared with other phenotypes. For example, at a cut-off of 10 µg Hb/g a cancer phenotype was 8.7 times more likely to have both samples positive compared with cases without neoplasia (OR=8.7; 95% CI=2.5 to 31.1; p=0.001). For all other phenotypes, positive concordance was always highest at a cut-off of 10 µg Hb/g and fell significantly at the higher cut-offs (figure 3, p<0.001 for each phenotype) except for cancer (p=0.89).

Sensitivity for advanced neoplasia according to test parameters

Figure 4 shows sensitivity for advanced neoplasia across a wide range of cut-offs for 1-sample and 2-sample tests. Sensitivity progressively rose as the cut-off value was lowered and was significantly higher for the 2-sample context at each cut-off. The rate ratios for advanced neoplasia (of sensitivity relative to those at 10 µg Hb/g)
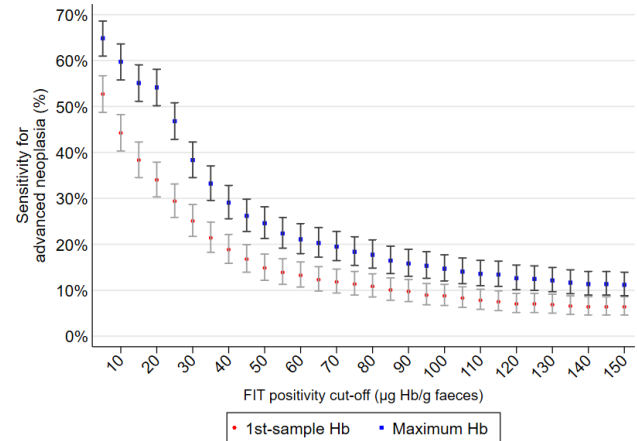
Figure 4 Sensitivity for advanced neoplasia across a wide range of cut-offs for 1-sample (first sample) and 2-sample (maximum of the two) tests in the cases where colonoscopy was performed within 365 days of the 2-sample FIT (regardless of FIT result). Bars show 95% CIs. FIT,faecal immunochemical test; Hb, haemoglobin.

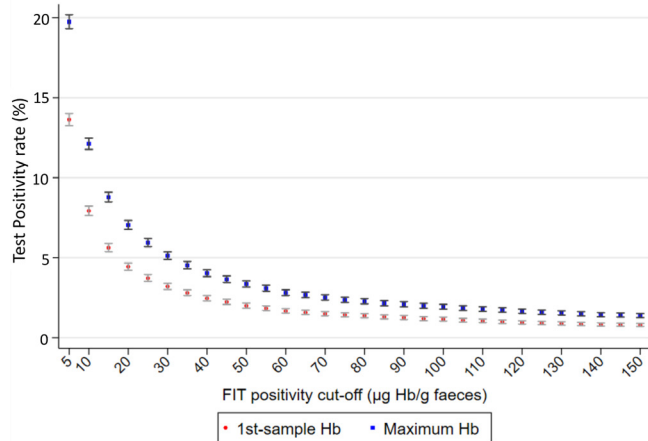for 20, 40 and 80 µg Hb/g were 0.77, 0.43 and 0.24, respectively (p<0.001 for each).

Actual sensitivities for cancer, advanced adenoma and advanced neoplasia at four selected cut-offs can be deduced for a 2-sample test by summing the percentages for positive concordance and discordant results shown in figure 3. For example, sensitivity for advanced adenoma was almost 60% at cut-off 10 µg Hb/g and fell to approximately 17% at cut-off 80 µg Hb/g.

Population test positivity rate according to test parameters

Figure 5 shows that population test positivity rates varied between 2% and 20% across a wide range of cut-offs for 1-sample and 2-sample contexts in all 32 413 instances where a 2-sample FIT was returned. At any given cut-off, test positivity was higher for 2-sample FIT.

Choosing test variables when sensitivity is the goal

The respective test cut-offs that would need to be applied to 1-sample and 2-sample tests to return an equivalent sensitivity for advanced neoplasia, were modelled together with the related population test positivity rate (table 2). For every given sensitivity, the f-Hb cut-off was higher for 2-sample FIT, due to the lack of concordance in positivity between samples and the extra opportunity to detect a lesion. It is also apparent that the population test positivity rate (colonoscopy workload) differed between sample contexts but the degree of difference was dependent on the desired sensitivity. At sensitivities ≥35%, the colonoscopy workload was higher with 1-sample compared with 2-sample FIT.

Figure 3 Test result concordance (agreement between the first and second samples) at selected cut-offs for the main phenotypes. AdvAd, advanced adenoma; Adv Neopl, advanced neoplasia; Non-neopl, non-neoplastic.

Figure 5 Population test positivity rate across a wide range of cut-offs for 1-sample (first sample) and 2-sample (maximum of the two) tests in all 32 413 instances where a 2-sample FIT was returned. Bars show 95% CIs. FIT, faecal immunochemical test; Hb, haemoglobin.

Choosing test variables when controlling colonoscopy workload is the goal

The respective cut-off values that would need to be applied to 1-sample and 2-sample tests to return equivalent colonoscopy workloads with 1-sample and 2-sample tests were modelled, together with respective sensitivities

for advanced neoplasia (table 3). For every given workload, the test cut-off was higher for 2-sample FIT. Sensitivity rates did not differ when seeking workloads in the range of 2% to 5%, but when workloads of 6% and above were addressed, sensitivity was higher with 2-sample FIT.

DISCUSSION

When establishing a CRC screening programme, many decisions need to be made to ensure that this multistep process succeeds. Reduction of the incidence of CRC is dependent on and achieved by removal of adenomas.[4–6 34] Effective detection of targeted neoplastic lesions along with feasibility of workloads are key considerations when deciding on the screening strategy. This study, in exploring agreement in f-Hb between samples and the relationship between sensitivity for advanced neoplasia and colonoscopy workload demonstrated that when setting programme goals for those using a quantitative FIT, both parameters of sample number as well as f-Hb cut-off for test positivity warrant consideration. Considering both adds additional flexibility to that provided by cut-off adjustment alone.

This study demonstrated that agreement in f-Hb between samples was poor, with wide variation seen between first and second samples. While such variation has been reported in the literature[16 35] the degree of

Table 2 Estimated test positivity cut-offs that must be applied to 1-sample and 2-sample tests to return an equivalent sensitivity for advanced neoplasia, together with respective population positivity rates (colonoscopy workloads)

Sensitivity for advanced neoplasia	Cut-off achieving sensitivity equivalence (n=626, µg Hb/g faeces)		Population test-positivity rate* (n=32 413)	
	1-sample	**2-sample†**	**1-sample**	**2-sample†**
55%	4.0	16.4	16.35% to 5298 15.9%–16.8%	8.18% to 2652 7.9%–8.5%
50%	6.0	22.4	11.74% to 3806 11.4%–12.1%	6.45% to 2091 6.2%–6.7%
45%	9.2	25.6	8.5% to 2755 8.2%–8.8%	5.83% to 1889 5.6%–6.1%
40%	12.6	28.4	6.57% to 2130 6.3%–6.8%	5.39% to 1748 5.1%–5.6%
35%	18.4	32.2	4.73% to 1532 4.5%–5.0%	4.86% to 1575 4.6%–5.1%
30%	24.2	38.8	3.82% to 1239 3.6%–4.0%	4.18% to 1354 4.0%–4.44%
25%	29.8	48.0	3.22% to 1045 3.0%–3.4%	3.46% to 1120 3.3%–3.7%
20%	37.6	68.6	2.63% to 852 2.5%–2.8%	2.57% to 832 2.4%–2.7%
15%	49.2	97.2	2.02% to 656 1.9%–2.2%	1.97% to 639 1.8%–2.1%
10%	83.8	157.2	1.30% to 422 1.2%–1.4%	1.30% to 422 1.2%–1.4%

*Positivity rate, n and 95% CI
†Maximum value of the two samples.
Hb, haemoglobin.

Table 3 Estimated positivity cut-offs for one and two sample contexts which give an equivalent population test positivity rate, together with respective sensitivities for advanced neoplasia

Population test-positivity rate	Cut-off achieving population test positivity rate equivalence (µg Hb/g faeces)		Sensitivity (%, for advanced neoplasia, 95% CI) for sample context at that cut-off (n=626)	
	1-sample	2-sample*	1-sample	2-sample*
2%	49.8	94.2	14.9% (n=93) 12.0%–17.7%	15.3% (n=96) 12.6%–18.4%
3%	32.4	56.2	22.8% (n=143) 19.3%–26.0%	21.9% (n=137) 18.9%–25.5%
4%	22.4	40.2	31.5% (n=197) 28.0%–35.4%	29.1% (n=182) 25.4%–32.6%
5%	17.0	30.6	36.3% (n=227) 32.3%–40.0%	37.2% (n=233) 33.3%–41.0%
6%	13.8	24.6	39.1% (n=245) 35.1%–42.9%	47.3% (n=296) 43.2%–51.1%
7%	11.6	20.0	41.4% (n=259) 37.3%–45.2%	54.3% (n=340) 50.2%–58.1%
8%	9.8	16.6	44.4% (n=278) 40.3%–48.2%	55.1% (n=345) 51.8%–58.9%
9%	8.4	14.4	45.8% (n=287) 41.7%–49.7%	55.8% (n=349) 51.6%–59.5%
10%	7.4	12.8	46.8% (n=293) 42.7%–50.6%	56.7% (n=355) 52.7%–60.6%

*Maximum value of the two samples.

variation and its effect on test positivity when adjusting cut-off, especially for advanced adenomas, has remained unclear. Because f-Hb concentrations for advanced neoplasia, especially advanced adenomas, often fell within the range of f-Hb cut-offs chosen in practice, FIT sensitivity for these lesions was compromised by the collection of one sample compared with two samples. Consequently, concordance between samples for a positive result was dependent on cut-off for positivity.

Our findings are consistent with previous observations in the literature showing that f-Hb (a reflection of the actual amount of blood loss) is dependent on neoplastic phenotype, being highest for cancer followed by advanced adenoma states.[2 7 11 24 26] Consequently and again in agreement with earlier reports,[2 7 11 24] we observed that sensitivity for neoplasia was increased by lowering the cut-off and/or increasing sample number, but in doing so the associated colonoscopy workload increased. But we found that such an inverse relationship is not a simple one when simultaneously but independently adjusting both variables. Concordance for test positivity between two samples was low, especially for advanced adenomas and especially when the cut-off was raised. Therefore, the proportion of advanced adenomas detected, and of cancer to a lesser extent, is critically dependent on the choice of each variable since detection is compromised if only one stool is to be sampled.

In a screening population, the proportion of tests that are positive determines the colonoscopy workload; aiming for a higher sensitivity results in a higher overall test positivity rate. We found that the workload rises progressively as the number of samples increases (unless the cut-off is raised) or as the cut-off is lowered (unless the number of samples is reduced). The highest colonoscopy workload was seen at the lowest cut-off while simultaneously using a 2-sample test because testing two samples reduces the chance that an f-Hb above the cut-off will be missed. The colonoscopy workload varied from 2% to 10% (in the cut-off range of 7.4 to 94.2 µg Hb/g) depending on the sample number/cut-off combination; respective sensitivities for advanced neoplasia varied from approximately 15% to 57%. In matching 1-sample and 2-sample FIT thresholds to achieve the same sensitivity, colonoscopy workload was significantly less with 2-sample FIT when the sensitivity goal exceeded 35%.

When a screening programme requires feasible management of workloads, different test cut-offs have been applied around the world.[13 14 36–38] A number of countries, including the Netherlands, UK and New Zealand,[13 36–38] have been cautious in undertaking programmes and the accepted colonoscopy workload sits at the lower levels of the wide range shown in table 3. For example, the Dutch programme moved from an initial cut-off of 15 µg Hb/g in 1-sample to a higher cut-off of 47 µg Hb/g due to the unmanageable workload.[14] Similarly, the screening programme in New Zealand increased from 15 µg Hb/g to 40 µg Hb/g.[13] When constraining workloads to this degree, our findings show that there is no sensitivity benefit in moving to a 2-sample approach (even when raising the cut-off to achieve equivalent colonoscopy rate).

On the other hand, where it is desired to maintain detection of advanced adenomas and where a colonoscopy workload of >5% is feasible, our findings show that a 2-sample test requires fewer colonoscopies for an equivalent sensitivity. In matching both parameters for equivalent colonoscopy workloads, sensitivity was similar in the workload range of 2% to 5%, but where higher workloads were feasible, better sensitivity was achieved with 2-sample FIT. Consequently, deciding which test variables to use is not simply a matter of managing feasibility of colonoscopic workload by adjusting the cut-off if the same number or additional lesions can be detected with equivalent or less colonoscopic workload. Programmes that have a goal of good detection of advanced adenomas should consider the 2-sample approach. Three-dimensional modelling of the relationship between colonoscopic effort (number needed to colonoscope to detect one case with advanced neoplasia), advanced neoplasia detection rate and colonoscopy workload in conditions aiming for best detection of advanced neoplasia showed that the 2-sample approach requires the least effort.[39] Our findings also align with those of a French modelling study where they found that the number needed to screen to detect a lesion was lower for the 2-sample test.[9] They observed similar positivity rate of 1.6% for 1-sample FIT at cut-off 68 µg Hb/g and 2-sample FIT at 116 µg Hb/g. Our findings might differ from those of a recent Dutch screening study conducted over four rounds where they found that workloads were higher with a 2-sample test but they did not adjust the cut-off for the 2-sample test and lesion detection was not assessed in all cases regardless of FIT result.[26]

Relative cost-effectiveness of 1-sample *versus* 2-sample testing has been considered in several studies. The cost per case detected in a FIT-screening study where only positive cases proceeded to colonoscopy was shown to be lowest for a 2-sample test compared with 1-sample and 3-sample tests.[40] In comparing cost-effectiveness of FIT screening by 1-sample and 2-sample methods by probabilistic modelling, both one and two samples were cost-effective and the per person cost varied by just over one dollar due to the provision of the extra collection tube in the 2-sample context.[41] A microsimulation model was used to compare costs (life years gained) drawing data from FIT screening studies in the Netherlands.[8] A 2-sample test provided additional life years gained but at additional cost. That study, and others, did not, however, consider costs and resultant colonoscopy workloads when the variables of cut-off and sample number were independently adjusted to achieve equivalent sensitivities. Modelling was also undertaken in populations when only those with a positive FIT were colonoscoped, meaning that the cost of missing adenomas was not available. In the current study, there were a large number of cases with neoplasia, especially advanced adenoma states, and sensitivity for neoplasia was estimated in cases who all underwent colonoscopy after completion of a 2-sample quantitative FIT, regardless of the test result. Our findings could now be incorporated into cost-effectiveness

modelling studies. While we accepted cases where FIT was performed anytime in the 12 months preceding colonoscopy and so assumed that the lesions found at colonoscopy were present at the time of taking the test, it is accepted that the natural history of colorectal neoplasia is of the order of a decade.[42]

While the population was not an average-risk naïve screening population there is no obvious reason to think that the principles which have emerged would be irrelevant to a typical screening population. The number of cancers in this study population was only 21 out of the 3499 meeting inclusion criteria for full modelling. As some in the study had previously undergone polypectomy, we would expect there to be a small number with cancer. The proportion with advanced adenomas, despite some colonoscopies being brought forward because of a positive FIT, were comparable to that of the entire population, and within 3% of the rate seen by others in a similar programme.[43] In addition, our FIT positivity rate for the 32 413 completed tests was similar to that reported in Scandinavian[44] and Australian screening populations.[45] The former reported positivity rates of 7.2% and 4.2% for a 2-sample and 1-sample FIT respectively at cut-off 20 ug/g, compared with our findings of 7.0% and 4.4%. Similarly high test positivity rates at a low cut-off were observed in a screening study in Israel.[24]

In the current study, we have not been able to consider differences in population participation based on the collection of just one rather than two samples. In a Dutch population-based randomised controlled trial comparing 1-sample and 2-sample FIT, the difference in participation rates was just 0.2%.[46] They also found that when seeking a higher detection rate for neoplasia, the 2-sample method was more efficient. In a US study, there was a small (4%) difference in participation rates between a 1-sample and 2-sample FIT but the actual stool-sampling device was very different between the two brands of FIT.[47] In other words, there is no conclusive evidence to support that participation in screening is improved when requiring just one sample.

In conclusion, variation in f-Hb between samples is such that both positivity cut-off and sample number are, together, crucial determinants of test sensitivity for advanced neoplasia (especially advanced adenomas) and for colonoscopy workload. If a programme goal includes reduction of cancer incidence, then detection of advanced adenomas is important. As test-positivity cut-offs applied in practice fall within the range of f-Hb observed in cases with advanced adenomas and cancers, their detection, especially of advanced adenomas, is compromised when sampling only one stool and especially when raising the cut-off as well. While the lower sensitivity of a 1-sample test can be addressed by choosing a lower cut-off than would be used for a 2-sample test, the relationship between these test parameters and outcomes is complex. If the sensitivity goal for advanced neoplasia is >35%, then colonoscopy workload is significantly less using two samples. Thus a 2-sample approach seems

preferable where programme detection of advanced adenomas has a priority alongside that of cancer.

Author affiliations
[1]Cancer Research, Flinders Health and Medical Research Institute, Flinders University, Adelaide, South Australia, Australia
[2]Biostatistics, Flinders Prevention, Promotion and Primary Health Care, General Practice, Flinders University, Adelaide, South Australia, Australia
[3]Flinders Centre for Innovation in Cancer, Flinders Medical Centre, Bedford Park, South Australia, Australia

ORCID iD
Graeme P Young http://orcid.org/0000-0001-9458-8383

REFERENCES

1. Young GP, Rabeneck L, Winawer SJ. The global paradigm shift in screening for colorectal cancer. *Gastroenterology* 2019;156:843–51.
2. Young GP, Symonds EL, Allison JE, *et al*. Advances in fecal occult blood tests: the fit revolution. *Dig Dis Sci* 2015;60:609–22.
3. Schreuders EH, Ruco A, Rabeneck L, *et al*. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015;64:1637–49.
4. Cole SR, Tucker GR, Osborne JM, *et al*. Shift to earlier stage at diagnosis as a consequence of the National bowel cancer screening program. *Med J Aust* 2013;198:327–30.
5. Zorzi M, Hassan C, Capodaglio G, *et al*. Long-term performance of colorectal cancerscreening programmes based on the faecal immunochemical test. *Gut* 2018;67:2124–30.
6. Chiu H-M, Chen SL-S, Yen AM-F, *et al*. Effectiveness of fecal immunochemical testing in reducing colorectal cancer mortality from the one million Taiwanese screening program. *Cancer* 2015;121:3221–9.
7. Young GP, Cole S. New stool screening tests for colorectal cancer. *Digestion* 2007;76:26–33.
8. Goede SL, van Roon AHC, Reijerink JCIY, *et al*. Cost-effectiveness of one versus two sample faecal immunochemical testing for colorectal cancer screening. *Gut* 2013;62:727–34.
9. Raginel T, Puvinel J, Ferrand O, *et al*. A population-based comparison of immunochemical fecal occult blood tests for colorectal cancer screening. *Gastroenterology* 2013;144:918–25.
10. Symonds EL, Fraser RJL, Young GP. FIT for purpose: enhanced applications for faecal immunochemical tests. *J Lab Precis Med* 2018;3:28.
11. Selby K, Levine EH, Doan C, *et al*. Effect of sex, age, and positivity threshold on fecal immunochemical test accuracy: a systematic review and meta-analysis. *Gastroenterology* 2019;157:1494–505.
12. Duffy MJ, van Rossum LGM, van Turenhout ST, *et al*. Use of faecal markers in screening for colorectal neoplasia: a European group on tumor markers position paper. *Int J Cancer* 2011;128:3–11.
13. New Zealand Ministry of Health. Bowel Screening - from pilot to national programme, 2017. Available: https://www.health.govt.nz/news-media/news-items/bowel-screening-pilot-national-programme
14. Toes-Zoutendijk E, van Leerdam ME, Dekker E, *et al*. Real-time monitoring of results during first year of Dutch colorectal cancer screening program and optimization by altering fecal immunochemical test cut-off levels. *Gastroenterology* 2017;152:767–75.
15. Macrae FA, St John DJ. Relationship between patterns of bleeding and Hemoccult sensitivity in patients with colorectal cancers or adenomas. *Gastroenterology* 1982;82:891–8.
16. Farrands PA, Hardcastle JD. Accuracy of occult blood tests over a six-day period. *Clin Oncol* 1983;9:217–25.
17. St. John DJB, Young GP, Alexeyeff MA, *et al*. Evaluation of new occult blood tests for detection of colorectal neoplasia. *Gastroenterology* 1993;104:1661–8.
18. Allison JE, Tekawa IS, Ransom LJ, *et al*. A comparison of fecal occult-blood tests for colorectal-cancer screening. *N Engl J Med* 1996;334:155–60.
19. Saito H, Yoshida Y. Mass screening: Japanese perspective. In: Young GP, Rozen P, Levin B, eds. *Prevention and early Dectection of colorectal cancer*. London: W. B. Saunders Company Ltd, 1996: 301–11.
20. Saito H. Issues in colorectal cancer screening in the Asian Pacific area. *Asian Pac J Cancer Prev* 2000;1:77–83.
21. Ee H, St John J. The National bowel cancer screening program: time to achieve its potential to save lives. *Public Health Res Pract* 2019;29. doi:10.17061/phrp2921915. [Epub ahead of print: 31 Jul 2019].
22. Mlakar DN, Bric TK, Škrjanec AL, *et al*. Interval cancers after negative immunochemical test compared to screen and non-responders' detected cancers in Slovenian colorectal cancer screening programme. *Radiol Oncol* 2018;52:413–21.
23. Young GP, Cole SR. Which fecal occult blood test is best to screen for colorectal cancer? *Nat Clin Pract Gastroenterol Hepatol* 2009;6:140–1.
24. Levi Z, Rozen P, Hazazi R, *et al*. A quantitative immunochemical fecal occult blood test for colorectal neoplasia. *Ann Intern Med* 2007;146:244–55.
25. Symonds EL, Pedersen SK, Baker RT, *et al*. A blood test for methylated BCAT1 and IKZF1 vs. a fecal immunochemical test for detection of colorectal neoplasia. *Clin Transl Gastroenterol* 2016;7:e137.
26. Schreuders EH, Grobbee EJ, Nieuwenburg SAV, *et al*. Multiple rounds of one sample versus two sample faecal immunochemical test-based colorectal cancer screening: a population-based study. *Lancet Gastroenterol Hepatol* 2019;4:622–31.
27. Bampton PA, Sandford JJ, Young GP. Achieving long-term compliance with colonoscopic surveillance guidelines for patients at increased risk of colorectal cancer in Australia. *Int J Clin Pract* 2007;61:510–3.
28. Symonds EL, Simpson K, Coats M, *et al*. A nurse-led model at public academic hospitals maintains high adherence to colorectal cancer surveillance guidelines. *Med J Aust* 2018;208:492–6.
29. Lane JM, Chow E, Young GP, *et al*. Interval fecal immunochemical testing in a colonoscopic surveillance program speeds detection of colorectal neoplasia. *Gastroenterology* 2010;139:1918–26.
30. Cancer Council Australia Surveillance Colonoscopy Guidelines Working Party. Clinical practice guidelines for surveillance colonoscopy: cancer Council Australia, Sydney, 2020. Available: https://wiki.cancer.org.au/australiawiki/index.php?oldid=208067, 2019
31. The National Health and Medical Research Council tARCaUA. *National statement on ethical conduct in human research 2007 (updated 2018)*, 2018.
32. Symonds EL, Osborne JM, Cole SR, *et al*. Factors affecting faecal immunochemical test positive rates: demographic, pathological, behavioural and environmental variables. *J Med Screen* 2015;22:187–93.
33. Bampton PA, Sandford JJ, Cole SR, *et al*. Interval faecal occult blood testing in a colonoscopy based screening programme detects additional pathology. *Gut* 2005;54:803–6.
34. Elmunzer BJ, Hayward RA, Schoenfeld PS, *et al*. Effect of flexible sigmoidoscopy-based screening on incidence and mortality of colorectal cancer: a systematic review and meta-analysis of randomized controlled trials. *PLoS Med* 2012;9:e1001352.
35. Hisamichi S, Fukao A, Fujii Y, *et al*. Mass screening for colorectal cancer in Japan. *Cancer Detect Prev* 1991;15:351–6.
36. Steele RJC, McDonald PJ, Digby J, *et al*. Clinical outcomes using a faecal immunochemical test for haemoglobin as a first-line test in a

national programme constrained by colonoscopy capacity. *United European Gastroenterol J* 2013;1:198–205.

37 Bowel Cancer UK. NHS England confirm new screening test to be introduced in Angland in 2018, 2017. Available: https://www.bowelcanceruk.org.uk/news-and-blogs/news/nhs-england-confirm-new-screening-test-to-be-introduced-in-england-in-2018/

38 Wales PH. Bowel screening Wales – important information on the new screening test, 2019. Available: http://www.wales.nhs.uk/sitesplus/888/news/50390

39 Young GP, Woodman R, FLI A, *et al*. Both sample number and test positivity threshold determine colonoscopy efficiency in colorectal cancer screening with quantitative fit. *Gastroenterology*. In Press 2020.

40 Yamamoto M, Nakama H. Cost-Effectiveness analysis of immunochemical occult blood screening for colorectal cancer among three fecal sampling methods. *Hepatogastroenterology* 2000;47:396–9.

41 Smith DH, O'Keeffe Rosetti M, Mosen DM, *et al*. Balancing adherence and expense: the cost-effectiveness of two-sample vs One-Sample fecal immunochemical test. *Popul Health Manag* 2019;22:83–9.

42 Winawer SJ. Natural history of colorectal cancer. *Am J Med* 1999;106:3–6. discussion 50S-51S.

43 Mangas-Sanjuan C, Santana E, Cubiella J, *et al*. Variation in colonoscopy performance measures according to procedure indication. *Clin Gastroenterol Hepatol* 2020;18:1216–23.

44 Ribbing Wilén H, Blom J, Höijer J, *et al*. Fecal immunochemical test in cancer screening - colonoscopy outcome in FIT positives and negatives. *Scand J Gastroenterol* 2019;54:303–10.

45 Lew J-B, St John DJB, Xu X-M, *et al*. Long-Term evaluation of benefits, harms, and cost-effectiveness of the National bowel cancer screening program in Australia: a modelling study. *Lancet Public Health* 2017;2:e331–40.

46 van Roon AHC, Wilschut JA, Hol L, *et al*. Diagnostic yield improves with collection of 2 samples in fecal immunochemical test screening without affecting attendance. *Clin Gastroenterol Hepatol* 2011;9:333–9.

47 Chubak J, Bogart A, Fuller S, *et al*. Uptake and positive predictive value of fecal occult blood tests: a randomized controlled trial. *Prev Med* 2013;57:671–8.