

Keywords: colorectal cancer screening; cost effectiveness; bowel cancer; sensitivity; specificity

Guaiac faecal occult blood test performance at initial and repeat screens in the English Bowel Cancer Screening Programme

B Kearns^{*1}, S Whyte¹, J Chilcott¹ and J Patnick²

¹School of Health and Related Research, University of Sheffield, Sheffield S1 4DA, UK and ²Public Health England, Sheffield S10 3TH, UK

Background: In many countries, screening for colorectal cancer (CRC) relies on repeat testing using the guaiac faecal occult blood test (gFOBT). This study aimed to compare gFOBT performance measures between initial and repeat screens.

Methods: Data on screening uptake and outcomes from the English Bowel Cancer Screening Programme (BCSP) for the years 2008 and 2011 were used. An existing CRC natural history model was used to estimate gFOBT sensitivity and specificity, and the cost-effectiveness of different screening strategies.

Results: The gFOBT sensitivity for CRC was estimated to decrease from 27.35% at the initial screen to 20.22% at the repeat screen. Decreases were also observed for the positive predictive value (8.4–7.2%) and detection rate for CRC (0.19–0.14%). Assuming equal performance measures for both the initial and repeat screens led to an overestimate of the cost effectiveness of gFOBT screening compared with the other screening modalities.

Conclusions: Performance measures for gFOBT screening were generally lower in the repeat screen compared with the initial screen. Screening for CRC using gFOBT is likely to be cost-effective; however, the use of different screening modalities may result in additional benefits. Future economic evaluations of gFOBT should not assume equal sensitivities between screening rounds.

Guaiac faecal occult blood tests (gFOBTs) are currently used to screen for colorectal cancer (CRC) within England via the Bowel Cancer Screening Programme (BCSP). This programme began roll-out in 2006 employing biennial screening. Initially people aged between 60 and 69 were eligible for screening, with the upper age-limit later extended to 74 years. Within the United Kingdom, CRC is the fourth most common form of cancer and the second most common form of cancer-related mortality (Cancer Research UK, 2013), with only half of all patients diagnosed with CRC surviving for more than 5 years (National Cancer Intelligence Network, 2013). Population-based screening for CRC is estimated to be cost-effective (Whyte *et al*, 2012) and has the potential to reduce CRC-related mortality due to earlier detection (and hence treatment) of the disease (Hewitson *et al*, 2008). Results from the first 1 million BCSP tests suggest that the programme is set to reduce CRC-related mortality by 16% (Logan *et al*, 2012).

There are several different modalities for CRC screening including: gFOBT, faecal immunochemical test for haemoglobin (FIT), flexible sigmoidoscopy (FS) and colonoscopy. The choice between screening modalities is determined by several factors: feasibility (availability of resources, clinicians and infrastructure to deliver the test to the population), acceptability (the willingness of the population to take part), effectiveness (determined by uptake, test sensitivity and specificity) and costs (including test administration and follow-up, along with treatment costs). Of these factors two key drivers are the specificity and sensitivity, which indicate how accurately subjects with and without the disease are classified by the screening test. Accurate estimates of sensitivity and specificity are difficult to obtain and they can vary considerably according to the population and the method of test administration (Kraemer, 1985; Burch *et al*, 2007; Leeftang *et al*, 2009).

*Correspondence: B Kearns; Email: b.kearns@sheffield.ac.uk

Received 6 May 2014; revised 23 July 2014; accepted 26 July 2014; published online 2 September 2014

© 2014 Cancer Research UK. All rights reserved 0007–0920/14

Initial screening refers to the first time a screening test is applied to an individual, future applications of the screening test to the same individual are referred to as repeat screens. An understanding of how test performance varies between initial and repeat screens is important to enable a robust comparison of screening strategies, which involve repeated screening. Data from a number of European countries suggest that, for screening programmes involving gFOBT, the positive predictive value of screening for both cancers and for adenomas decreases between the initial and repeat screening rounds (Steele *et al*, 2010; Benson *et al*, 2012). However, it is unknown whether these decreases between initial and repeat screens relate to changes in the disease spectrum of the population (the distribution of cancers and adenomas, as only a nonrandom subset of those receiving an initial screen will receive a repeat screen) or to changes in the test characteristics of sensitivity and specificity between initial and repeat screening. Most studies on the cost effectiveness of repeated gFOBT screening, including the 2011 re-appraisal of the options for CRC screening in England (Whyte *et al*, 2012), assume that gFOBT sensitivity and specificity do not vary between screening rounds (Frazier *et al*, 2000; O'Leary *et al*, 2004; Song *et al*, 2004; Wu *et al*, 2006; Macafee *et al*, 2008; Tsoi *et al*, 2008; Heitman *et al*, 2010).

During the 2011 reappraisal, an unexpected difference between model predictions and observed data was noted for the repeat screen (Whyte *et al*, 2012). This difference motivated the current study, which had two aims. The first was to investigate if gFOBT performance measures, including sensitivity and specificity, differed between initial and repeat screens. The second aim was to assess the implications of any differences on the effectiveness and cost-effectiveness of the different screening modalities that may be used within the BCSP.

MATERIALS AND METHODS

Data used. Data were available from the BCSP for all persons invited to screening in England between the years 2006 and 2011. This comprises 9.8 million invitations to attend an initial screen of which 4.6 million people were screened, and 2.0 million invitations to attend a repeat screen of which 1.8 million people were screened. A breakdown of the screening data by year is presented in Table 1. Of those who were screened, data were available on the outcome of the gFOBT screen, and, of those who were referred for follow-up, data were available on whether any cancers or adenomas were detected. It was hypothesised that during the initial roll-out of the data there may have been a learning effect while the capacity to deliver follow-up colonoscopy at a national level was established, and that this may have affected the representativeness of the data. Because of this, and because of the small number of repeat screens

in these 2 years, data for the first 2 years (2006 and 2007) were excluded. The remaining data (from 2008 onwards) were used to estimate detection rates and positive predictive values (PPV) for CRC, low-risk and high-risk adenomas. Uptake and positivity rates were also calculated. Where possible, performance measures were based on those defined by the International Colorectal Cancer Screening Network (Benson *et al*, 2012). Adenoma risk was based on British Society of Gastroenterology guidelines (Cairns *et al*, 2010). Definitions for the measures used in this study are provided in Box 1.

Colorectal cancer natural history model. An existing mathematical model that represents the natural history of CRC along with screening pathways was used to estimate sensitivity and specificity and to also generate cost-effectiveness estimates. This model was used for the 2011 re-appraisal of the options for screening in England (Whyte *et al*, 2012), and synthesised incidence and prevalence data and screening data from the BCSP, a UK trial of FS (Atkin *et al*, 2010) and data on FIT from Italy (Castiglione *et al*, 2002). The model incorporates the benefits of screening from both a reduction in incidence via the removal and surveillance of adenomas, and a reduction in CRC mortality via earlier diagnosis of CRC.

Estimates of the natural history of CRC progression from normal colorectal epithelium through the development of adenomas, CRC and eventual mortality were obtained, as were the performance measures of sensitivity and specificity for each of the screening modalities. Health-care resource use and costs, along with the impact on patient's health-related quality of life were evaluated for a range of screening options. Full technical details have been previously published (Whyte *et al*, 2011a, 2012) and details relevant to this study are described below.

The model is a cohort-Markov model built in Microsoft Excel, which simulates both the natural history of CRC and the possible pathways of screening participants, as shown in Figure 1. Accurate observational data for natural history transition rates and for sensitivity and specificity were not available, so Bayesian calibration methods were used to jointly estimate these (Whyte *et al*, 2011b). These unobserved parameters were linked to observed outcomes via the CRC natural history model, the calibration methods searched over possible parameter values, and selected those which best fit the observed data. Screening outcomes for an individual depend on both their underlying natural history and the effectiveness of the screening modality. This effectiveness was measured by a number of performance measures, which consists of sensitivity, specificity, PPV, uptake, positivity, and detection rates for CRC, high-risk adenomas and low-risk adenomas. The model allowed estimates of sensitivity to vary by disease severity and estimates of specificity to vary by age.

Table 1. Breakdown of screening data available from the English Bowel Cancer Screening Programme, by year

Year	Initial screen			Repeat screen		
	Invited	Screened	Uptake %	Invited	Screened	Uptake %
2006	54 779	28 051	51	<5	0	0
2007	740 890	380 659	51	788	303	38
2008	1 956 674	1 020 008	52	16 147	13 981	87
2009	2 238 854	1 141 565	51	242 540	208 049	86
2010	2 483 968	1 147 347	46	813 549	710 632	87
2011	2 374 449	909 389	38	970 350	854 437	88
Total	9 849 614	4 627 019	47	2 043 374	1 787 402	87
2008 to 2011	9 053 945	4 218 309	47	2 042 586	1 787 099	87

For this analysis, the model was adapted and re-calibrated to allow performance measures to vary between initial and repeat screens. It was assumed that all repeat screens had equal performance measures. Variations in performance measures by age were also explored, to assess if any between-round differences were specific to any age groups.

Statistical methods. To convey an assessment of uncertainty, 95% confidence intervals were calculated. All of the intervals are based on the Wilson method (Newcombe, 2012). The differences between the initial and repeat screens were tested for statistical significance using the two-proportion Z-test. Evidence of an age gradient was assessed by fitting linear regression models (weighted by the number of observations). A P-value less than 0.05 was deemed to be statistically significant.

Effectiveness and cost effectiveness. The results from the adapted and re-calibrated model were used to estimate the cost effectiveness of various CRC screening options. This study considered the same screening options as the 2011 re-appraisal, which includes the modalities gFOBT, FIT, FSIG and combinations. Effectiveness was measured in terms of CRC cases avoided and lives saved. Cost effectiveness was measured as the incremental cost per incremental quality-adjusted life year (QALY) gained. The analysis took the perspective of the NHS and personal social services in England. Both costs and QALYs were discounted at a rate of 3.5% per year, as recommended by the National Institute for Health and Care Excellence (2013). A lifetime horizon was employed.

RESULTS

Between 2008 and 2011, a total of 9 053 945 first invitations for an initial screen were sent out, along with 2 042 586 first invitations for a repeat screen. Performance measures for the initial and repeat screens are displayed in Table 2. All of the differences were statistically significant ($P < 0.001$). Uptake, positivity rate, detection rate and PPVs are based on observed data from the BCSP, while sensitivity and specificity are estimated via the model calibration process. For sensitivity and specificity, the estimated values were used to derive the denominator value.

Performance measures derived directly from observed data. During the initial screen a total of 4 218 309 useable kits were returned, producing an uptake of 46.59% (95% confidence interval 46.56–46.62%). This figure increased to 87.49% (87.45–87.54%) for the repeat screen. Of those who responded to initial screening using gFOBT, 93 355 had a positive result requiring further investigation, giving a positivity rate of 2.21% (2.20–2.23%) for the initial screen, which dropped to 1.97% (1.96–1.20%) for the repeat screen.

Detection rates for both CRC and high-risk adenomas were both higher for the initial screen than for the repeat screen.

Box 1 Definitions used within this study

- Initial screen: The first time a person is screened.
- Repeat screen: A screen in a person who has previously been screened.
- Small adenoma: an adenoma that is less than 1cm.
- Low-risk adenomas health state: people with one or two small adenomas.
- High-risk adenomas health state: people with at least three small adenomas, or at least one adenoma greater than 1cm.
- Uptake: the number of results obtained for a test divided by the number of invitations.
- Positivity: the number of people with a positive test results (those persons referred to follow-up colonoscopy) divided by the total number of test results.
- Adenoma (or CRC) detection rates: the number of people with adenomas (or CRCs) detected by a test, divided by the tested population.
- Positive predictive value for adenomas (or CRC): the proportion of positive test results that were shown to be an adenoma (or CRC) at colonoscopy.
- Sensitivity to adenomas (or CRC): the probability that a test gives a positive result amongst people with adenomas (or CRC).
- Specificity to adenomas (or CRC): the probability that a test gives a negative result amongst people without adenomas (or CRC).

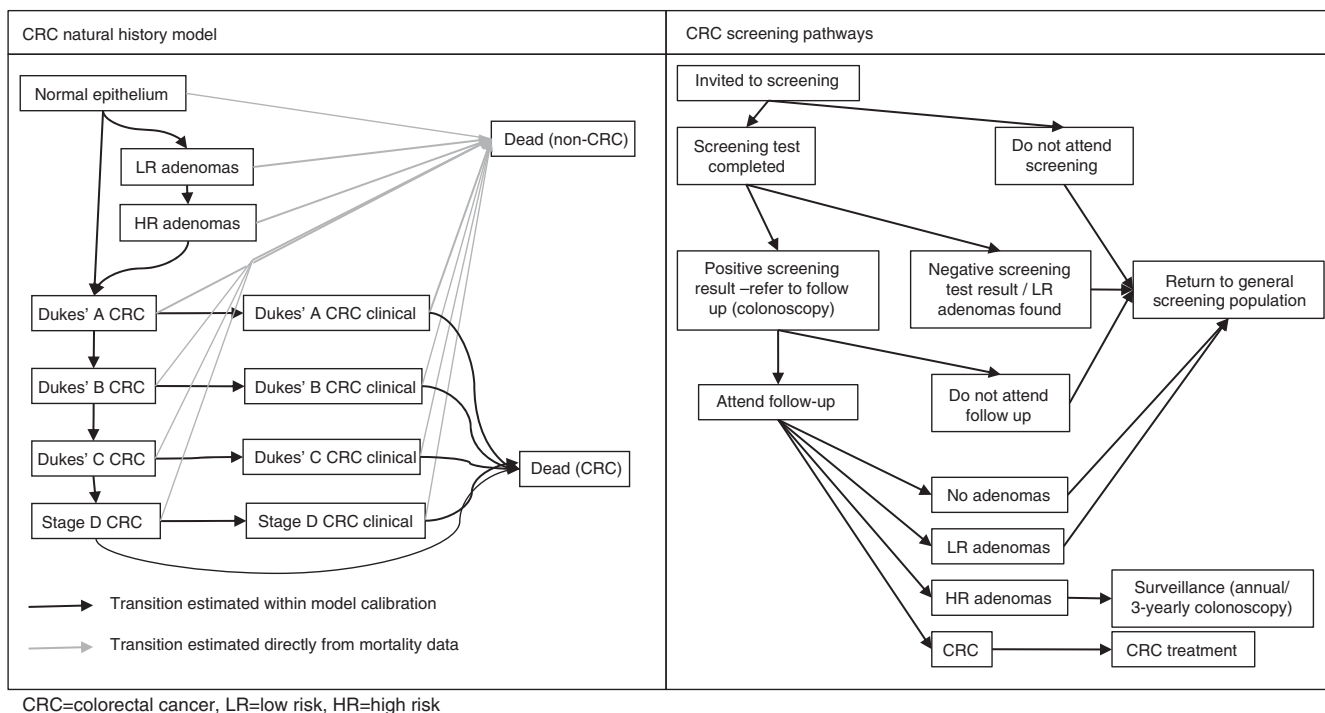


Figure 1. Schematic of the natural history model for CRC and the screening pathways, as used for this study. Reproduced with permission from Whyte et al (2012).

Table 2. Observed and estimated performance measures (values given as percentages and the corresponding denominator (*n*) value), along with a 95% confidence interval

	Initial screen % (n) (95% CI)	Repeat screen % (n) (95% CI)
Performance measures derived directly from observed data		
Uptake	46.591 (9 053 945) (46.558 to 46.623)	87.492 (2 042 586) (87.447 to 87.537)
Positivity rate	2.213 (4 218 309) (2.199 to 2.227)	1.975 (1 787 099) (1.955 to 1.995)
Detection rates		
Colorectal cancer	0.185 (4 218 309) (0.181 to 0.189)	0.141 (1 787 099) (0.136 to 0.147)
High-risk adenomas	0.542 (4 218 309) (0.535 to 0.549)	0.393 (1 787 099) (0.384 to 0.402)
Low-risk adenomas	0.308 (4 218 309) (0.302 to 0.313)	0.316 (1 787 099) (0.307 to 0.324)
Positive predictive value		
Colorectal cancer	8.361 (93 355) (8.185 to 8.540)	7.157 (35 295) (6.893 to 7.430)
High-risk adenomas	24.494 (93 355) (24.219 to 24.771)	19.909 (35 295) (19.496 to 20.329)
Low-risk adenomas	13.896 (93 355) (13.676 to 14.120)	15.977 (35 295) (15.598 to 16.363)
Performance measures estimated using mathematical model		
Sensitivity		
Colorectal cancer	27.349 (28 538) (26.835 to 27.869)	20.220 (12 492) (19.525 to 20.934)
High-risk adenomas	13.508 (169 280) (13.346 to 13.671)	9.463 (74 254) (9.255 to 9.676)
Low-risk adenomas	1.114 (1 164 489) (1.095 to 1.133)	1.084 (520 078) (1.056 to 1.113)
Specificity		
Age 65	97.922 (238 419) (97.864 to 97.979)	97.314 (139 728) (97.228 to 97.398)
Age 70	97.773 (18 693) (97.552 to 97.975)	97.133 (99 120) (97.027 to 97.235)
Abbreviations: CI = confidence interval; CRC = colorectal cancer.		

For CRC, the detection rate fell by over a quarter (in relative terms) from 0.185 (0.181–0.189%) to 0.141% (0.136–0.147%), a similar (relative) decrease was observed for high-risk adenomas, from 0.542 (0.535–0.549%) to 0.393% (0.384–0.402%).

Colorectal cancer was identified in 7805 of respondents with a positive gFOBT during the initial screen, giving a PPV of 8.36% (8.19–8.54%), which fell to 7.16% (6.89–7.43%) during the repeat screen. The PPVs for high-risk adenoma also decreased, from 24.49 (24.22–24.77%) to 19.91% (19.50–20.33%).

Performance measures estimated using mathematical model.

The sensitivity of gFOBT to detecting CRC during the initial screen was estimated to be 27.35% (26.84–27.87%), dropping to 20.22% (19.53–20.93%) for the repeat screen. A similar pattern is seen for high-risk adenomas; sensitivity drops from 13.51% (13.35–13.67%) during the initial screen to 9.46% (9.26–9.68%) during the repeat screen.

Estimated specificity values varied by age, for every increase in age by 1 year, specificity decreased by 0.03% in the initial screen and by 0.04% in the repeat screen. Estimated specificities for the ages 65 and 70 years are presented as examples. During the initial screen the estimates were 97.92% (97.86–97.98%) for age 65 and 97.77% (97.55–97.98%) for age 70 years. For the repeat screen, these values were 97.31% (97.23–97.40%) and 97.13% (97.03–97.24%), respectively.

Variations in gFOBT performance measures by age at initial and repeat screen. Variation by age in uptake and detection rates are displayed in Figure 2, while differences in positivity rate and PPVs are displayed in Figure 3.

The results presented in Figures 2 and 3 show that the (statistically significant) differences between initial and repeat screen are observed for all of the age groups, and are not caused by differences amongst age subgroups. The figures also emphasise that there were relatively small numbers of individuals aged over 70

years in the screening programme (for both rounds) and a small number of individuals aged 61 years receiving a repeat screen, as reflected by the wide confidence intervals among these ages.

There was no evidence of a statistically significant age gradient for uptake rates or for low-risk adenoma PPVs at either screen. All of the other performance measures had statistically significant age gradients for both screens. These conclusions remained the same when individuals aged above 70 years were excluded from the analysis.

Impact on cost effectiveness. Estimates of cost effectiveness for 13 different screening options are presented in Table 3. Results are discounted per person over their lifetime and ordered by increasing costs. A strategy of no screening is the cheapest, but also results in the fewest QALYs. Of the strategies involving only a single modality, those involving FIT dominate those involving either gFOBT or FS. The hybrid strategy of FS at age 55 years followed by biennial FIT between the ages of 60 and 74 years is the most effective (generates the most QALYs), but is also more expensive than any of the strategies involving FIT alone. Uncertainty in these results was assessed using probabilistic sensitivity analyses (as described in the previous re-appraisal (Whyte *et al*, 2012)), and was found to be small.

Compared with no screening, a strategy of biennial gFOBT screening between the years of 60 and 69 is estimated to generate an additional 0.005 QALYs, at an additional cost of £12, resulting in an incremental cost-effectiveness ratio (ICER) of £2420. Extending the upper age-limit to 74 increases this ICER to £2900. Both of these values are comfortably below the threshold of £20 000 often used to determine cost effectiveness at a national level. By not modelling differences in gFOBT test characteristics at repeat screens, the cost effectiveness of gFOBT screening is overestimated. There is no difference in average costs, but for biennial gFOBT at age 60–69 the average QALYs accrued increases

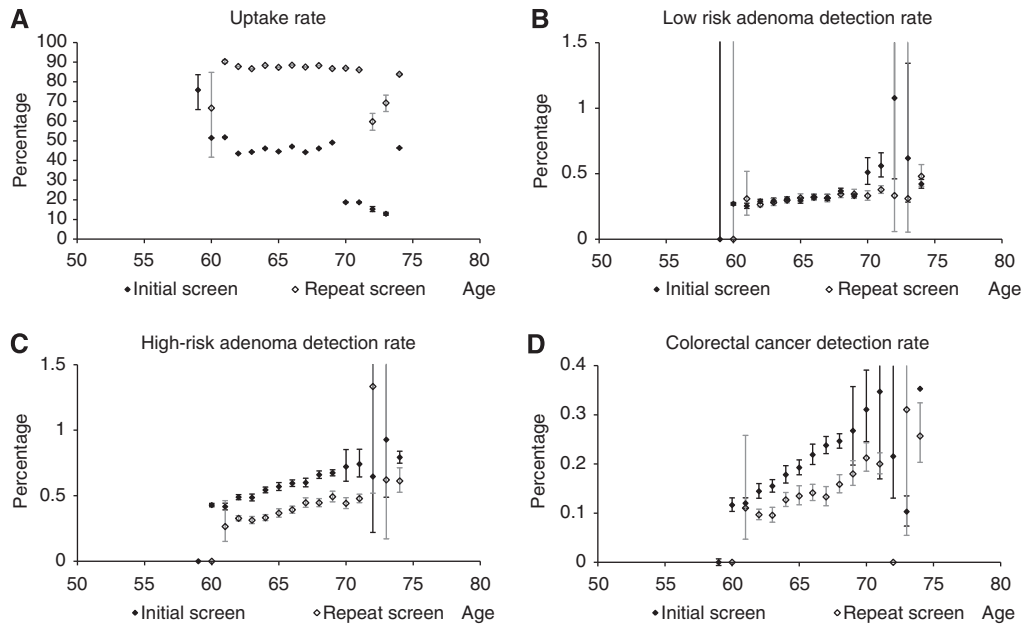


Figure 2. (A) Variation in uptake rate by age, (B) Variation in low-risk adenoma detection rate by age, (C) Variation in high-risk adenoma detection rate by age, (D) Variation in colorectal cancer detection rate by age. For all, 95% confidence intervals are displayed, although for some data points these are not noticeable due to their short width.

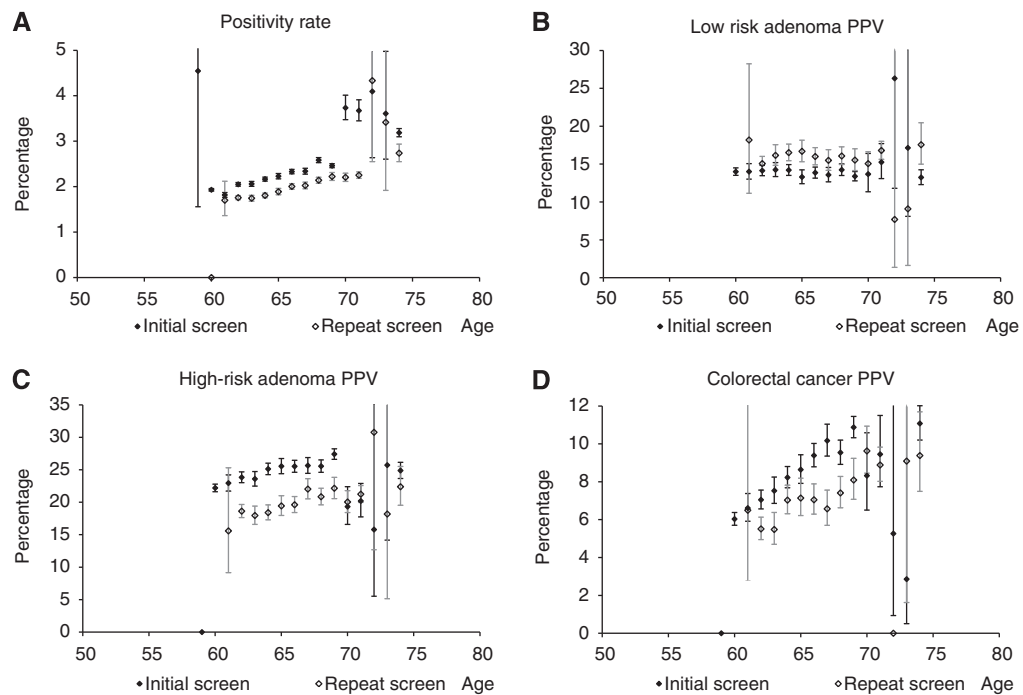


Figure 3. (A) Variations in positivity rate by age, (B) variations in low-risk adenoma positive predictive value by age, (C) variations in high-risk adenoma positive predictive value by age, (D) variations in colorectal cancer positive predictive value by age. For all, 95% confidence intervals are displayed, although for some data points these are not noticeable due to their short width.

to 19.060, resulting in an ICER of £1922 compared with no screening. For biennial gFOBT at age 60–74 the average QALYs accrued increases to 19.062, resulting in an ICER of £2213 compared with no screening.

DISCUSSION

The results of this study show the performance of the English gFOBT BCSP against a number of key performance measures for

both initial and repeat screens. There was an increase in uptake rates for the repeat screen compared with the initial screen. However, for CRC and for high-risk adenomas, the performance measures of sensitivity, specificity, PPV and detection rate were all lower in the repeat screen compared with the initial screen. There was little difference in these performance measures for low-risk adenomas.

For cancer screening programmes, uptake rates have been identified as key measures (Steele *et al*, 2009). An overall uptake of 47% for the initial screen was observed in this study, which compares with an uptake of 87% for the repeat screen. A lower

Table 3. Comparison of the cost effectiveness of different screening modalities

Strategy	Costs	QALYs	ICER	Comparator
No screening	£78	19.054		
gFOBT at 60–69 (biennial)	£90	19.059	Ext. Dom.	FIT at 60, 65, 70 years
gFOBT at 60–74 (biennial)	£95	19.060	Ext. Dom.	FIT at 60, 65, 70 years
FIT at 60, 65, 70 years	£96	19.062	£2150	No screening
FIT at 60–69 (biennial)	£107	19.066	£3002	FIT at 60, 65, 70 years
FIT at 60–74 (biennial)	£120	19.069	£5658	FIT at 60–69 (biennial)
FS at age 55	£182	19.060	Dominated	FIT at 60–74 (biennial)
FS age 55, gFOBT 66–74 (biennial)	£192	19.063	Dominated	FIT at 60–74 (biennial)
FS age 55, FIT 60, 65, 70	£200	19.067	Dominated	FIT at 60–74 (biennial)
FS age 55, FIT 66–74 (biennial)	£208	19.067	Dominated	FIT at 60–74 (biennial)
FS age 55, FIT 60–74 (biennial)	£224	19.072	£27,868	
FS age 55, FIT 56–74 (biennial)	£236	19.054	Dominated	FIT at 60–74 (biennial)
FS age 55, 65	£253	19.065	Dominated	FIT at 60–74 (biennial)

Abbreviations: gFOBT = guaiac faecal occult blood test; FIT: faecal immunochemical test for haemoglobin, FS: flexible sigmoidoscopy. ICER: Incremental cost-effectiveness ratio. Ext. Dom.: extensively dominated. Intervention A is dominated by intervention B if the latter is cheaper and produces more QALYs. Intervention A is dominated by intervention B if the latter is cheaper and has a lower ICER. Currently used strategy is shown in italics.

uptake rate for initial screening (compared with repeat screening) is to be expected, as initial screening is offered to the entire eligible population, whereas repeat screening is only offered to people who previously accepted screening, and so are likely to accept a repeat screen (Janda *et al*, 2010). Uptake rates for initial screening will also be affected if there exists a subgroup of ‘persistent refusers’ in the eligible population.

For both CRC and high-risk adenomas, there was a drop in detection rates, PPVs and sensitivities between the initial and repeat screens. Similar decreases have been observed for gFOBT screening programmes across the world (Benson *et al*, 2012). A limited number of studies have estimated how the sensitivity of gFOBT varies between screening rounds. The UK (Nottingham) pilot (Moss *et al*, 1999) (which used biennial screening), using the proportional incidence method (Day, 1985), reported CRC sensitivities of 62.7% for the first screen and 50.0% for the rescreen. A method based on the number of observed interval cancers was applied to data from the Province of Florence to give CRC sensitivities of 38% for the first screen and 62% for subsequent screens (Zappa *et al*, 2001). There is a large degree of variation in both the estimates of CRC sensitivity and the methods used to derive them. The use of interval cancers to estimate sensitivity has two main limitations; first, some interval cancers would not have been cancers at the time of the screen and second, not all undetected cancers will be diagnosed within any given interval. The proportional incidence method alleviates the first limitation by replacing interval cancers with the incidence of cancers in the absence of screening. However, it does not explicitly consider the second limitation. The values presented here are based on explicitly modelling the natural history of CRC in the absence of screening, and so address both of these limitations. The CRC sensitivities reported in this study (27% for the first screen and 20% for the repeat screen) are lower than the estimates from the UK pilot and the Province of Florence estimates. This is likely to be due to the slow-growth rate of early-stage cancers, which may not present symptomatically for a number of years. For example, one study found an overall FOBT sensitivity to CRC of 12.9%, which fell to 6.7% for TNM Stage I cancers (Imperiale *et al*, 2004). It should also be noted that the sensitivity estimates reported for this study relate to an individual screen, gFOBT screening is carried out on a biennial basis within the BCSP, so the overall programme sensitivity will be much higher.

Decreases in detection rates and PPV at the repeat screen may be due to the initial screen reducing the prevalence of CRC and high-risk adenomas among people attending for repeat screening, without reducing the false-positive rate (Steele *et al*, 2010). This effect should not lead to changes in sensitivity and specificity as these two measures are theoretically independent of underlying disease prevalence. However, it has been shown in practice that changes in disease spectrum can affect sensitivity and specificity (Kraemer, 1985; Leeftang *et al*, 2009). Differences in CRC and adenoma prevalence at the initial and repeat screens were included within the natural history model, hence these differences alone did not explain the decrease in gFOBT test performance. It is possible that other changes in disease spectrum that lead to the decrease in PPVs and detection rates observed in this study may also have led to the estimated decreases in sensitivity and specificity. In addition to a lower CRC prevalence at repeat screening, there may also be other disease spectrum differences. For example, completed gFOBT samples are classified as abnormal or normal depending on the presence or absence of blood in the faecal samples. It may be hypothesised that patients with heavy bleeding are more likely to be detected in the first round, which would lead to a decrease in detection rates for repeat screening.

The cost effectiveness of different screening strategies for CRC was re-assessed using different test characteristics for the initial and repeat rounds of gFOBT. The results indicated that screening with FIT (possibly preceded by FS) was to be preferred to screening with either gFOBT, FS or a combination of the two. It should be stressed that the choice of screening modality depends on more than just the cost-effectiveness results, as other criteria such as the feasibility and acceptability of each modality must also be considered. While the BCSP currently uses biennial gFOBT screening, screening with FIT (Logan *et al*, 2012) is being evaluated, and FS is currently being piloted.

There are a number of existing economic evaluations of the cost effectiveness of gFOBT screening for CRC (Lansdorp-Vogelaar *et al*, 2011), covering areas such as the United States (Frazier *et al*, 2000), United Kingdom (Macafee *et al*, 2008), Australia (O’Leary *et al*, 2004), France (Berchi *et al*, 2004) and Asia (Tsoi *et al*, 2008). There are differences between these economic evaluations due to variations in screening programmes between countries. These differences include the use of private health care in the United States of America and the use of different screening algorithms,

including different test thresholds for referral (Benson *et al*, 2012). However, it is important to note that all of these existing economic evaluations have assumed that the performance measures of gFOBT sensitivity and specificity are the same across screening rounds. The results of this study indicate that this may not be the optimal approach, and that more realistic estimates of the effectiveness of gFOBT would be obtained by modelling separate performance measures across screening rounds.

Owing to a lack of available data, when calculating the cost effectiveness results for this study it was assumed that all repeat screens had equal test characteristics (such as sensitivity). In addition, it was also assumed that performance measures for FIT screening were the same for the initial and repeat screens.

There are a number of strengths to this study. The BCSP provides a large volume of data covering all of England, allowing for accurate estimates of the observed performance indicators (uptake, detection rates, and PPVs). Use of a mathematical model allows for the comparison of screening programmes based on different evidence sources within a consistent framework, and permits the estimation of unobservable quantities such as the natural history of CRC in the absence of screening. The mathematical model previously assumed equal test characteristics between screening rounds. This structural uncertainty was explored in this study by allowing the screening test characteristics to vary between initial and repeat screens. Comparing different model structures helps to provide a greater understanding of the natural history of CRC and how this interacts with the screening test characteristics.

In summary, this study has shown that both observed and estimated performance measures show a decrease when comparing initial and repeat screening. The results presented here indicate that screening for CRC is still likely to represent a cost-effective strategy, but also that there is room for further development, for example by exploring the use of different screening modalities. The results from this study suggest that future evaluations of gFOBT screening should not assume equal test characteristics (such as sensitivity and specificity) between screening rounds.

ACKNOWLEDGEMENTS

The authors acknowledge the help of Claire Nickerson, who provided the data for this work.

REFERENCES

- Atkin WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JM, Parkin DM, Wardle J, Duffy SW, Cuzick J (2010) Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* **375**(9726): 1624–1633.
- Benson VS, Atkin WS, Green J, Nadel MR, Patnick J, Smith RA, Villain P (2012) Toward standardizing and reporting colorectal cancer screening indicators on an international level: the International Colorectal Cancer Screening Network. *Int J Cancer* **130**(12): 2961–2973.
- Berchi C, Bouvier V, Réaud JM, Launoy G (2004) Cost-effectiveness analysis of two strategies for mass screening for colorectal cancer in France. *Health Econ* **13**(3): 227–238.
- Burch JA, Soares-Weiser St K, John DJB, Duffy S, Smith S, Kleijnen J, Westwood M (2007) Diagnostic accuracy of faecal occult blood tests used in screening for colorectal cancer: a systematic review. *J Med Screen* **14**(3): 132–137.
- Cairns SR, Scholefield JH, Steele RJ, Dunlop MG, Thomas HJ, Evans GD, Eaden JA, Rutter MD, Atkin WP, Saunders BP (2010) Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). *Gut* **59**(5): 666–689.
- Cancer Research UK (2013) Cancer Statistics: Key Facts: Bowel Cancer April 2013. http://publications.cancerresearchuk.org/downloads/Product/CS_KF_BOWEL.pdf.
- Castiglione G, Grazzini G, Miccinesi G, Rubeca T, Sani C, Turco P, Zappa M (2002) Basic variables at different positivity thresholds of a quantitative immunochemical test for faecal occult blood. *J Med Screen* **9**(3): 99–103.
- Day NE (1985) Estimating the sensitivity of a screening test. *J Epidemiol Community Health* **39**(4): 364–366.
- Frazier AL, Colditz GA, Fuchs CS, Kuntz KM (2000) Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA* **284**(15): 1954–1961.
- Heitman SJ, Hilsden RJ, Au F, Dowden S, Manns BJ (2010) Colorectal cancer screening for average-risk North Americans: an economic evaluation. *PLoS Med* **7**(11): e1000370.
- Hewitson P, Glasziou P, Watson E, Towler B, Irwig L (2008) Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am J Gastroenterol* **103**(6): 1541–1549.
- Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME (2004) Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* **351**(26): 2704–2714.
- Janda M, Hughes KL, Auster JF, Leggett BA, Newman BM (2010) Repeat participation in colorectal cancer screening utilizing fecal occult blood testing: a community-based project in a rural setting. *J Gastroenterol Hepatol* **25**(10): 1661–1667.
- Kraemer HC (1985) The robustness of common measures of 2×2 association to bias due to misclassifications. *Am Stat* **39**(4): 286–290.
- Lansdorp-Vogelaar I, Knudsen AB, Brenner H (2011) Cost-effectiveness of colorectal cancer screening. *Epidemiol Rev* **33**(1): 88–100.
- Leeftang MM, Bossuyt PM, Irwig L (2009) Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* **62**(1): 5–12.
- Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C (2012) Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut* **61**(10): 1439–1446.
- Macafee DAL, Waller M, Whynes DK, Moss S, Scholefield JH (2008) Population screening for colorectal cancer: the implications of an ageing population. *Br J Cancer* **99**(12): 1991–2000.
- Moss SM, Hardcastle JD, Coleman DA, Robinson MH, Rodrigues VC (1999) Interval cancers in a randomized controlled trial of screening for colorectal cancer using a faecal occult blood test. *Int J Epidemiol* **28**(3): 386–390.
- National Cancer Intelligence Network (2013) Colorectal cancer survival by stage. http://www.ncin.org.uk/publications/data_briefings/colorectal_cancer_survival_by_stage.
- National Institute for Health and Care Excellence (2013) *Guide to the methods of technology appraisal*. <http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9>.
- Newcombe RG (2012) *Confidence intervals for proportions and related measures of effect size*. CRC Press: Boca Raton, FL, USA.
- O'Leary BA, Olynyk JK, Neville A, Platell CF (2004) Cost-effectiveness of colorectal cancer screening: comparison of community-based flexible sigmoidoscopy with fecal occult blood testing and colonoscopy. *J Gastroenterol Hepatol* **19**(1): 38–47.
- Song K, Fendrick AM, Ladabaum U (2004) Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. *Gastroenterology* **126**(5): 1270–1279.
- Steele RJC, Kostourou I, McClements P, Watling C, Libby G, Weller D, Brewster DH, Black R, Carey FA, Fraser C (2010) Effect of repeated invitations on uptake of colorectal cancer screening using faecal occult blood testing: analysis of prevalence and incidence screening. *BMJ* **341**: c5531.
- Steele RJ, McClements PL, Libby G, Black R, Morton C, Birrell J, Mowat NAG, Wilson JA, Kenicer M, Carey FA (2009) Results from the first three rounds of the Scottish demonstration pilot of FOBT screening for colorectal cancer. *Gut* **58**(4): 530–535.
- Tsoi KKF, Ng SSM, Leung MCM, Sung JY (2008) Cost-effectiveness analysis on screening for colorectal neoplasm and management of colorectal cancer in Asia. *Aliment Pharmacol Ther* **28**(3): 353–363.
- Whyte S, Chilcott J, Essat M, Stevens J, Wong R, Kalita N (2011a) Reappraisal of the options for colorectal cancer screening. Report for the NHS Bowel Cancer Screening Programme. http://www.screening.nhs.uk/policydb_download.php?doc=130.

- Whyte S, Chilcott J, Halloran S (2012) Reappraisal of the options for colorectal cancer screening in England. *Colorectal Disease* **14**(9): e547–e561.
- Whyte S, Walsh C, Chilcott J (2011b) Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Med Decis Making* **31**(4): 625–641.
- Wu GH, Wang YM, Yen AM, Wong JM, Lai HC, Warwick J, Chen TH (2006) Cost-effectiveness analysis of colorectal cancer screening with stool DNA testing in intermediate-incidence countries. *BMC Cancer* **6**(1): 136.
- Zappa M, Castiglione G, Paci E, Grazzini G, Rubeca T, Turco P, Crocetti E, Ciatto S (2001) Measuring interval cancers in population-based screening using different assays of fecal occult blood testing: the district of Florence experience. *Int J Cancer* **92**(1): 151–154.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.