
Article:
Thomas, C., Whyte, S., Kearns, B. et al. (1 more author) (2019) External validation of a colorectal cancer model against screening trial long term follow-up data. Value in Health, 22 (10). pp. 1154-1161. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2019.06.005

External Validation of a Colorectal Cancer Model against Screening Trial Long Term Follow-up Data

Running Title: Validation of a colorectal cancer screening model

Chloe Thomas PhD,

ScHARR,

University of Sheffield,

Sheffield,

United Kingdom.

Sophie Whyte PhD,

ScHARR,

University of Sheffield,

Sheffield,

United Kingdom.

Benjamin Kearns MSc,

ScHARR,

University of Sheffield,

Sheffield,

United Kingdom.

James B Chilcott MSc,

ScHARR,

University of Sheffield,

Sheffield,

United Kingdom.

Corresponding Author:

Chloe Thomas PhD

ScHARR

Regent Court

30 Regent Street

University of Sheffield

Sheffield S1 4DA

United Kingdom

c.thomas@sheffield.ac.uk

Phone +44 (0)114 2220785

Funding/Support: This study was supported by funding from the UK National Screening Committee.

Financial Disclosure: None reported

Summary: The ScHARR bowel cancer screening model can successfully replicate the results of long-term follow-up screening trials to within 95% confidence intervals.

Acknowledgements: We would like to acknowledge Wendy Atkin and Kate Wooldrage for their useful comments and advice around validation of the UKFSST. We would also like to thank Claire Nickerson, Matt Rutter, Neil Hawkes, Stephen Halloran, Alistair Grey, Jacqueline Murphy, Roland Valori, Dominic Blunt, Stephen Duffy and Anne Mackie and John Marshal of the UK NSC who all contributed invaluable advice, expertise and support during development of the version of the ScHARR bowel cancer screening model used for this analysis.

Word Count: 3,998

Number of Pages: 28

Number of Figures: 3

Number of Tables: 2

Appendix

Pages: 22

Figures: 1

Tables: 4

Abstract

Objectives

The ScHARR Bowel Cancer Screening Model has been used to make decisions about screening strategies in England. The objective of this study was to perform external validation of the ScHARR model against long-term follow-up data about colorectal cancer (CRC) incidence and mortality reductions due to screening, from the Nottingham Trial of guaiac Faecal Occult Blood Testing for CRC, and the UK Flexible Sigmoidoscopy Screening Trial.

Methods

The ScHARR model was adapted prior to validation to reflect the setting of each trial in terms of population characteristics, details of screening and surveillance programmes, uptake of screening and further investigations and study follow-up. The impact of using current versus historical CRC incidence and mortality data in the validation was also examined by carrying out a series of analyses in which historical data from different years was included in the model.

Results

The ScHARR model was able to predict CRC incidence and mortality rate/hazard ratios from both trials to well within the 95% confidence intervals in the observed data. Whilst it was less accurate in predicting absolute incidence and mortality rates, modelling historical incidence and mortality data enabled these predictions to be improved considerably.

Conclusions

The ScHARR model is able to replicate the long-term relative benefit from screening observed in two large-scale UK based screening trials and can therefore be considered to be an appropriate tool to facilitate decision making around the English bowel cancer screening programme.

Highlights

- ISPOR guidelines indicate that model credibility depends upon their external validation. The ScHARR Bowel Cancer Screening Model has been used to make decisions about screening in England, but its external validity has not previously been demonstrated.
- The ScHARR model closely predicts colorectal cancer incidence and mortality hazard ratios from two screening trials; the Nottingham Trial of guaiac Faecal Occult Blood Testing and the UK Flexible Sigmoidoscopy Screening Trial.
- Accuracy of predictions is improved if model parameters are adjusted to reflect historical changes in colorectal cancer incidence and mortality that have occurred since the trials were carried out.

Introduction

Screening for colorectal cancer (CRC) in England has been carried out over the past decade through the NHS Bowel Cancer Screening Programme (BCSP) [1]. Decisions around which screening strategies to implement have been informed by analyses using the ScHARR Bowel Cancer Screening Model. This is a cohort state transition model built in Excel that models the life experience of a cohort of individuals aged 30 from the general population of England, starting with normal epithelium, through to the development of adenomas, CRC and subsequent death (see supplementary material for further details). The cohort can undergo a vast variety of different CRC screening strategies, or no screening, according to pre-specified inputs [2,3]. Whilst much of the model can be populated using estimates from published data, cancer natural history and screening test characteristics are unobservable processes that cannot be modelled directly and instead require model calibration to select parameters that produce model predictions fitting the observed outcomes (see Whyte et al., 2011 [4]).

The credibility of a model depends upon it having undergone a thorough process of verification and validation, to minimise errors and ensure that the model reflects reality sufficiently to enable decision making [5]. This is particularly important where calibration has been carried out to estimate unknown parameters. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines on model transparency and validation state that there are five main types of validation that should be considered; face validity, internal validity, cross validity, external validity and predictive validity [6]. The ScHARR model has undergone an extensive process of internal validation as part of model development, whilst reviewing of model results by clinical experts has ensured their face validity. Cross validity has also been assessed as part of a recent project [3], through comparative analysis of results from the ScHARR model with those from a model built

independently by researchers from Oxford University [7]. All these processes are helpful for identifying errors, but do not assess how accurately the model can represent novel scenarios.

Data is now available for external validation of the ScHARR model. Over the past few years, long-term outcomes for CRC incidence and mortality from two large-scale UK CRC screening trials have been published; the Nottingham Trial of guaiac Faecal Occult Blood Testing (gFOBT) for CRC (20 year follow-up)[8]; and the UK Flexible Sigmoidoscopy Screening Trial (UKFSST) (11 and 17 year follow-up) [9,10]. The Nottingham gFOBT trial was conducted between 1981 and 1995 and randomised 152,850 individuals aged between 45 and 74 years to either biennial gFOBT screening or to control (no screening) groups [11], whilst the UKFSST randomised 170,432 individuals aged between 55 and 65 to one-off flexible sigmoidoscopy (FS) or to control, and took place between 1994 and 1999 [10]. Long-term follow-up of between 11 and 20 years indicates that both gFOBT and FS screening result in significant CRC mortality reductions, but that only FS screening has a significant impact in reducing CRC incidence.

In this analysis we show that the ScHARR model can successfully predict the results of long-term follow-up of the Nottingham gFOBT trial and the UKFSST, in particular the relative benefits of screening compared to no screening. We also demonstrate how the accuracy of model predictions is affected if model inputs that are known to have changed over time such as CRC incidence and mortality rates, and other cause mortality are adjusted before validation to reflect as closely as possible the historical situation in which the trial was conducted.

Methods

The background to the ScHARR model is summarised in the supplementary information and described in detail elsewhere [3]. In brief, it consists of a natural history model with 13 health states relating to normal epithelium, low risk adenoma, high risk adenoma, CRC (4 stages, clinical and preclinical) and death (from CRC and other causes) (supplementary Figure S1). Parameters for natural history transitions and symptomatic diagnosis were calibrated against CRC incidence data [12]; parameters for mortality were taken from survival data (for CRC mortality) and Office for National Statistics (ONS) Life Tables and Death certificate data (for other cause mortality) [13-15]; parameters for screening sensitivity and specificity were calculated using data from the English BCSP [16] combined with model calibrated prevalence, whilst other parameters were taken from the BCSP where possible, and from the literature. These are summarised in supplementary Table S4.

In order to improve the success of model validation against the Nottingham gFOBT and UKFSST long-term follow-up data [8-10], a series of adaptations to the ScHARR model were made in order to reflect the setting of each study as closely as possible. Note that none of these changes involved using the long-term follow-up data from either screening study as a parameter input or a calibration target, as the aim of the analysis was to externally validate the model, not internally validate it. Full explanation of model adaptations and limitations with respect to each study can be found in supplementary tables S1 and S2.

Population:

CRC incidence and mortality varies by population characteristics including age, gender and socioeconomic deprivation [17,18]. The model was adapted to reflect the age structure of individuals enrolled in the studies as closely as possible; a similar approach has been used for validations of other CRC models [19]. The cohort-level nature of the model meant that other

characteristics from the trials could not be included in the model. However, the gender breakdowns reported in both trials were similar to that in the English population, indicating that the model would be able to represent an average trial participant in terms of gender. It was also not possible to incorporate study exclusion criteria (typically those at high risk of disease or with comorbid conditions), which may mean that the modelled population was less healthy than the study population [20].

Previously screened individuals are known to be at lower risk of cancer than those who have never been screened [21]. Both studies were initiated prior to the implementation of screening programmes in England, therefore it was assumed in the model that no-one had previously been screened.

Screening programme:

Both studies gave details about the screening programme followed, including invitation and test protocol, type of screen performed, type and quality of further investigation for positive individuals, number of screening rounds, interval between screening round and age at which screening was offered. Screening programme details were modelled as closely as possible, but some aspects were not possible to replicate. For example, barium enema was used as an alternative follow-up investigation to colonoscopy in some cases in the Nottingham gFOBT trial [11], but this was not available within the current structure of the ScHARR model. Another complication of the Nottingham gFOBT trial that could not be replicated in the model was the modification of the re-invitation protocol eligibility criteria part-way through the study (Supplementary Table S1).

Information about detection rates was given in each study, enabling inference of relative sensitivity of the test compared to current screening within the BCSP. Detection rates and sensitivity are proportional if the underlying age-specific prevalence of CRC and adenomas is

the same. The method used for calculation of sensitivity from detection rates is described in more detail in the supplementary information. Detection rates in the UKFSS trial were higher than those currently seen within the BCSP for FS screening [1,22], so model sensitivity for FS (originally derived in the model through calibration against BCSP data) was increased proportionally as shown in supplementary Table S3.

Uptake and completion of screening and subsequent investigations:

Screening uptake was modified to reflect the information reported in each study. No information was given in the studies about uptake of further investigations following a positive test, so default values based on English BCSP data were used. Whilst uptake of screening and uptake of further investigations in those who test positive is known to vary by screening round, screening history and by personal characteristics [23-25]; and the Nottingham gFOBT trial did give information about uptake by screening round [11], this could not be incorporated within the structure of the model.

Surveillance:

Both studies indicated that individuals received surveillance following polypectomy, however no details were given relating to the criteria for being offered surveillance, stopping surveillance, or the surveillance method or frequency. It was therefore assumed that surveillance followed the current UK guidelines (see supplementary information for details).

Length of study follow-up:

Study follow-up varied between individuals enrolled in the studies, however it was not possible to incorporate this within the cohort structure of the model. Mean study follow-up time was therefore modelled for each study.

Historical and geographical setting:

Data such as other-cause mortality, CRC mortality and CRC incidence vary significantly over time and between different geographical settings. For example, current day CRC and all-cause mortality in the UK is much lower than it was historically [26], presumably due to advances in treatment and early diagnosis through screening. Historical life tables and death certificate data are available from the ONS [13,14], whilst CRUK has information about changes in age-standardised CRC incidence and mortality over time [26]. This was used to calculate incidence and mortality multipliers that were applied to the transition from normal epithelium to low risk adenoma in the model, in order to represent CRC incidence and mortality at historical time periods. This is described in more detail in the supplementary information.

The Nottingham gFOBT trial took place between 1981 and 1995 [8]. The impact of using current versus historical data in the model was tested by carrying out a series of analyses in which historical life tables and death certificate data, together with incidence and mortality multipliers from different years was used where available, to represent either current data, data from 1981 (the start of the trial), from 1995 (end of the screening period), or from 2005 (half-way between the end of the screening period and follow-up).

The UKFSS trial was carried out between 1996 and 1999 with follow-up outcomes published in 2010 and 2017 [9,10]. The impact of using current versus historical data was tested by carrying out a series of analyses in which historical life tables and death certificate data, together with incidence and mortality multipliers from different years was used where available, to represent either current data, data from 1996 (the start of the trial); from 2003 (halfway between the trial start and first follow-up), or from 2010 (first follow-up date).

Outcomes:

The model was run deterministically in order to provide estimates of study outcomes. For all analyses, model performance was validated against both relative outcomes (measured as

hazard ratios or rate ratios) and absolute outcomes (measured as rates per 100,000 person years). For the comparison against the UKFSST, validation against long-term data for both published time-points (11 years and 17 years) was carried out. In the basecase analyses, no changes were made to incidence and mortality data, and for the Nottingham trial validation it was assumed that 4 screening episodes were carried out. Deterministic sensitivity analysis was used to examine the impact of modelling historical incidence and mortality as described above, and to examine the impact of modelling either 3 or 5 screening episodes for the Nottingham trial validation.

Results

Validation against the Nottingham gFOBT trial

The Nottingham gFOBT trial 20 year follow-up analysis reported a significant reduction in certified CRC mortality following gFOBT screening compared to no screening, with a rate ratio of 0.91 (95% CI: 0.84 to 0.98) [8] (Table 1). A reduction in CRC incidence was also observed, with a rate ratio of 0.97 (95% CI: 0.91 to 1.03), but this was not significant. Using current day incidence and mortality data, and assuming four screening episodes per person (the basecase scenario), the model is able to replicate this reduction in CRC mortality due to gFOBT screening very accurately, but estimates a higher reduction in CRC incidence than observed in the trial, although results are well within the reported 95% confidence interval. Sensitivity analysis indicates that if the number of screening episodes is reduced to three per person, CRC incidence and mortality are slightly higher in the screening arm, whereas if the number of screening episodes is increased to five per person, CRC incidence and mortality are slightly reduced.

Whilst the model can accurately replicate rate ratios, absolute incidence and mortality rates are poorly replicated by the model using current day incidence and mortality data. The model tends to overestimate CRC incidence in both the control and screening arms by about 13%, but underestimate absolute CRC and all-cause mortality rates by up to one third (Table 1). These differences are corrected in the sensitivity analyses that use historical incidence and mortality data, with the results suggesting that absolute CRC incidence rate and absolute all-cause mortality in the Nottingham gFOBT trial might be best estimated using data from a period between 1981 and 1995, and absolute CRC mortality best estimated using data from a period between 1995 and 2005, to reflect increased mortality rates and reduced incidence of CRC in the past.

However, use of historical data does not appear to improve the estimate of the incidence and mortality rate ratios. Use of historical data results in overestimation of CRC mortality rate ratios, with the poorest estimates coming from the model using the oldest data (albeit still within 95% confidence intervals). It is important to note that the trial CRC mortality rates shown in Table 1 represent certified CRC mortality, whereas reported verified CRC mortality rates were higher (100 per 100,000 person years in the control arm and 91 per 100,000 person years in the screening arm). It is unclear which type of mortality rate is best represented by the model output as the model uses both death certificate (certified) and survival (verified) data.

Comparison of cumulative CRC incidence and mortality over time in the trial and in the model (using 1995 incidence and mortality data) is shown in Figure 1. The incidence curves indicate that the model is closely representing the increase in CRC incidence seen in the first few years following screening initiation. Cumulative CRC incidence starts to be reduced in the screening arm compared to the control arm at around 10 years following screening initiation in the model, whilst this appears to take place somewhere between year 7 and year 15 in the trial. The mortality curves indicate that in the trial, CRC mortality is reduced within 3-4 years of initiating screening compared to the control arm. However, this is not apparent in the model, where the two curves do not start to separate until after year 10. Unlike the trial, the model also predicts slightly higher CRC mortality in the screening arm in the years following screening initiation in all analyses undertaken (Figure 1 and data not shown).

Validation against the UKFSST

The UKFSST follow-up analysis reported a significant reduction in both CRC incidence and certified CRC mortality following FS screening compared to no screening, at both 11 and 17 year time-points [9,10]. The data suggests that CRC incidence reductions continue to accumulate

slightly beyond 11 years (hazard ratio at 11 years = 0.77 [95% CI: 0.70-0.84]; hazard ratio at 17 years = 0.74 [95% CI: 0.70-0.80]), whilst mortality reductions stay roughly constant between the two time points (HR at 11 years = 0.69 [95% CI: 0.59-0.82]; HR at 17 years = 0.70 [95% CI: 0.62-0.79]). Using current day incidence and mortality data (basecase scenario), these hazard ratios are replicated well within 95% confidence intervals using the model (Table 2). Whilst data from other time points is not available from the trial, the model predicts that the incidence reduction in the screening arm compared to the control arm peaks at 15 years following screening, whilst the CRC mortality reduction peaks at 16 years post screening (Figure 2). Absolute incidence and mortality rates per 100,000 person years are replicated reasonably well by the model using current day incidence and mortality data, with estimates for both tending to be slightly lower than observed data (Table 2). Use of historical incidence and mortality data in sensitivity analysis reduces the accuracy of model absolute incidence estimates for all modelled years, whilst CRC and all-cause absolute mortality estimates are too high using the oldest data and too low using the most recent data, suggesting using data from a period between 2003 and 2010 might produce the best estimates. Incidence has not substantially changed since the mid1990s when the UKFSST was initiated, which is likely to explain the lack of improvements produced through using historical incidence data.

Comparison of cumulative CRC incidence and mortality over time in the trial and in the model (using current day incidence and mortality data) is shown in Figure 3. The incidence curves indicate that the model is closely representing the increase in CRC incidence seen in the first few years following screening initiation. Cumulative CRC incidence starts to be reduced in the screening arm compared to the control arm at around six years following screening initiation in the model, whilst this appears to take place at year five in the trial. The mortality curves indicate that in the trial, CRC mortality is reduced soon after initiating

screening compared to the control arm. However, this is not apparent in the model, where the two curves do not start to separate until after year five. In common with the Nottingham gFOBT trial validation, the model predicts slightly higher CRC mortality in the screening arm in the early years following screening initiation in all analyses undertaken (Figure 3 and data not shown).

Discussion

Validation of models against external data is essential to ensure their credibility for decision making; however, considerable model adaptation may be required before a model can accurately replicate trial data. The analyses presented here indicate that following model adaptation, the ScHARR model is able to replicate the long-term relative benefit from screening observed in two large-scale UK based screening trials, with a level of accuracy that falls well within trial 95% confidence intervals [8-10]. Interestingly, the model is less able to replicate trial data for absolute mortality and incidence rates than it is for relative rates. However, the results suggest that this can be improved somewhat by approximating historical mortality and incidence data in the model, particularly for the Nottingham gFOBT trial that started in 1981, when both incidence and mortality were quite different from today. This shows that it is important to build the ability to use historical data into the model if accurate validation against long-term data is to be performed. It also implies that estimation of the effects of future screening programmes could potentially be made more accurate through projection of incidence and mortality trends into the future. Implementing the ability to include incidence and mortality data that vary over time would therefore improve the ScHARR model, both for validation and for model predictions.

There are limitations to accurate modelling of historical data that were encountered as part of this project. In particular, the lack of detailed historical CRC incidence and survival data by age and stage represent a general problem for model validation and prediction. The problem was overcome in this analysis by using mortality and incidence multipliers. For simplicity, incidence multipliers were applied on the first transition from normal epithelium to low risk adenoma. However, there is no evidence that the general increase in CRC incidence over time is due to a higher rate of this transition, and it is more likely that all transitions between health states show some historical variation. Applying the multiplier at this point implies that

incidence of adenomas has also increased over time, and therefore that screening may be of more benefit. A further limitation of this approach is that it ignores any differences in age or stage distribution of incidence and mortality between historical and current data. This could impact on the ability of the model to replicate trial data and therefore to be successfully validated if for example stage distribution at diagnosis has shifted significantly to earlier stages recently. The introduction of screening has also skewed incidence data over the past decade for the screening eligible age group, meaning that underlying changes in incidence that may have happened since screening began (e.g. due to changing lifestyles) cannot be captured.

Aside from the problems in using historical data, there are several other limitations that were encountered. Firstly, study populations often do not resemble the screened general population due to exclusion from participation of individuals with serious illnesses or at high risk of CRC due to symptoms or family history. This is likely to result in a trial population being healthier than the general population in a way that is not easily represented in a model, therefore having lower CRC incidence and mortality than expected, and be less likely to benefit from screening (called the healthy volunteer effect) [20]. It was reported by Atkin et al., (2010) that CRC incidence in the study control group was almost exactly as expected in the general population [10], but an equivalent comparison for CRC mortality was not given. Secondly, there may be complexities within trial screening pathways that a model cannot capture without significant restructuring. For example, individuals in the Nottingham gFOBT trial originally did not receive subsequent invitations to screening if they had not attended the first screening round – but this was later changed [11]. The model however assumes that everyone is re-invited for every eligible screening round. We now know from BCSP data that those who have not previously attended screening have a higher incidence of CRC and polyps than those who have previously attended screening [21], so this change in screening programme

would have impacted study results, although the magnitude of effect when compared with model results is likely to be small given a fixed number of screening episodes. Thirdly, in the past 35 years, colonoscopy quality is likely to have improved [27], potentially leading to greater detection and more successful removal of adenomas. All of these reasons could lead to overestimation of screening effectiveness.

Differences in model predictions and long-term follow-up data can be caused by structural model assumptions, which can highlight areas for future model improvements. For example, the tendency of the ScHARR model to underestimate CRC mortality reductions, particularly in the first couple of years, may be partly due to model assumptions that survival of screen-detected cancers by stage is identical to survival following opportunistic detection of cancer. In fact, recent data suggests that survival is much higher for screen-detected cancers, which cannot be entirely explained by the shift towards earlier stage at diagnosis [28,29]. The availability of detailed differential stage and age specific survival data for cancers detected through different diagnostic pathways will allow model predictions to be improved considerably.

Validation of CRC models against 11 year follow-up data from the UKFSST has previously been carried out by members of the Cancer Intervention and Surveillance Modelling Network (CISNET) [30], through which three independent CRC microsimulation models have been developed [31-33], each quite different structurally to the ScHARR model. In common with the ScHARR model, all three models underestimated CRC mortality reductions, whilst incidence reductions were underestimated by two models (one outside of 95% confidence intervals) and overestimated by the other [19]. Of all four models, the ScHARR model provides the closest estimates of incidence reduction and the second closest for CRC mortality reduction. This is likely to be due to the UK focus of the ScHARR model and the careful consideration around which adaptations should be carried out to attempt to reproduce the study setting.

Conclusion

The ScHARR model is able to replicate the long-term relative benefit from screening observed in two large-scale UK based screening trials; the Nottingham gFOBT trial and the UKFSST, with a level of accuracy that falls well within trial 95% confidence intervals. Accuracy of predictions is improved if model parameters are adjusted to reflect historical changes in colorectal cancer incidence and mortality that have occurred since the trials were carried out. Overall, this analysis indicates that the ScHARR Bowel Cancer Screening Model is an appropriate tool to facilitate decision making around the NHS BCSP.

Table 1: Validation of the ScHARR Bowel Cancer Screening Model against the 20 year follow-up of the Nottingham gFOBT trial [8] showing absolute rates per 100,000 person years and rate ratios for gFOBT screening versus no screening, using either current-day or historical CRC incidence, CRC mortality and all-cause mortality data in the model. For current-day results, the impact of altering the number of screening episodes (SE) between 3 and 5 per person is also shown.

	CRC Incidence		**CRC Mortality**		**All-Cause Mortality**	
	Control	***Screening***	***Control***	***Screening***	***Control***	***Screening***
Trial	182	176	88	80	3,130	3,140
Rate ratio (95% CI)	*0.97 (0.91-1.03)*		*0.91 (0.84-0.98)*		*1.00 (0.99-1.02)*	
Model: Current: 4 SE	207	196	72	64	2,088	2,076
Rate ratio	*0.95*		*0.91*		*1.00*	
Model: Current: 3 SE	207	198	72	66	2,088	2,078
Rate ratio	*0.96*		*0.92*		*1.00*	
Model: Current: 5 SE	207	194	72	63	2,088	2,075
Rate ratio	*0.94*		*0.89*		*1.00*	
Model: 1981: 4 SE	174	165	93	88	3,576	3,566
Rate ratio	*0.95*		*0.95*		*1.00*	
Model: 1995: 4 SE	195	185	93	87	3,012	3,001
Rate ratio	*0.95*		*0.94*		*1.00*	
Model: 2005: 4 SE	204	193	80	73	2,457	2,445
Rate ratio	*0.95*		*0.92*		*1.00*	
CRC = colorectal cancer; CI = confidence interval; SE = screening episodes						

Table 2: Validation of the ScHARR Bowel Cancer Screening Model against the UKFSS trial [9,10] showing absolute rates per 100,000 person years and hazard ratios for FS screening versus no screening, using either current-day or historical CRC incidence, CRC mortality and all-cause mortality data in the model, at either 11 years or 17 years.

	CRC Incidence		**CRC Mortality**		**All-Cause Mortality**	
	Control	***Screening***	***Control***	***Screening***	***Control***	***Screening***
11 Year Follow-up						
Trial	149	114	44	30	1,124	1,093
Hazard ratio (95% CI)	*0.77 (0.70-0.84)*		*0.69 (0.59-0.82)*		*0.97 (0.94-1.00)*	
Model: Current Day	145	112	38	28	917	902
Hazard ratio	*0.77*		*0.73*		*0.98*	
Model: 1996	139	108	51	42	1,488	1,475
Hazard ratio	*0.78*		*0.82*		*0.99*	
Model: 2003	139	108	45	36	1,201	1,188
Hazard ratio	*0.78*		*0.79*		*0.99*	
Model: 2010	145	112	37	26	976	962
Hazard ratio	*0.77*		*0.72*		*0.99*	
17 Year Follow-up						
Trial	184	137	56	39	1,483	1,472
Hazard ratio (95% CI)	*0.74 (0.70-0.80)*		*0.70 (0.62-0.79)*		*0.99 (0.97-1.01)*	

Model: Current Day	179	136	53	37	1,309	1,289
Hazard ratio	*0.76*		*0.69*		*0.98*	
Model: 1996	169	129	69	52	2,116	2,095
Hazard ratio	*0.76*		*0.75*		*0.99*	
Model: 2003	171	130	62	45	1,732	1,712
Hazard ratio	*0.76*		*0.73*		*0.99*	
Model: 2010	178	136	51	35	1,394	1,374
Hazard ratio	*0.76*		*0.68*		*0.99*	
CRC = colorectal cancer; CI = confidence interval						

Figure 1: Estimates of cumulative colorectal cancer (CRC) incidence (top) and mortality (bottom) for gFOBT screening versus no screening from the Nottingham gFOBT trial study [8] (solid lines) and the model (dashed lines) using 1995 incidence and mortality data with 20 years of follow-up from first screen.

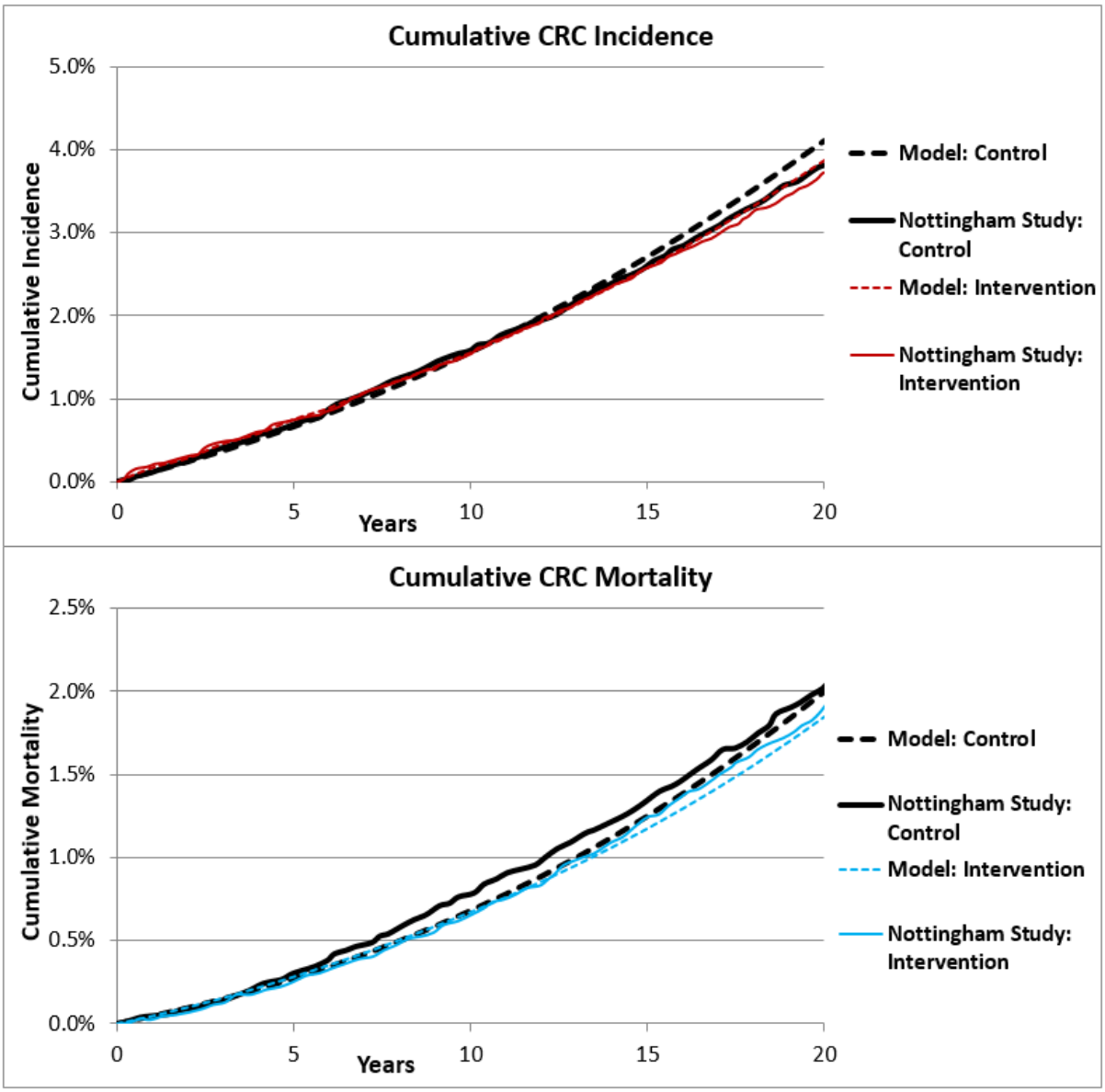


Figure 2: Model projected hazard ratios for every year of a lifetime horizon for FS screening versus no screening using current day incidence and mortality data. Observed data from the UK Flexible Sigmoidoscopy Screening Trial (UKFSST) [9] is plotted as point estimates with error bars indicating 95% confidence intervals. The graph indicates that peak mortality reductions in the screening arm would be expected 16 years post screening, whilst peak reduction in incidence would be expected to occur 15 years after screening (dotted vertical lines).

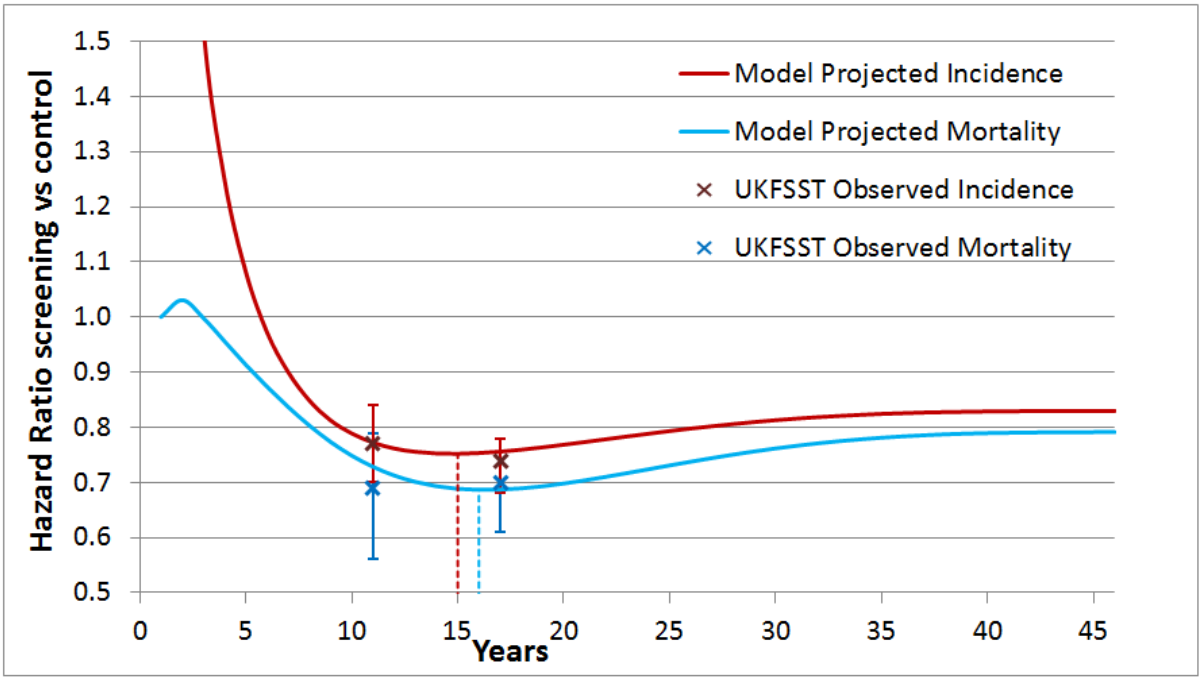
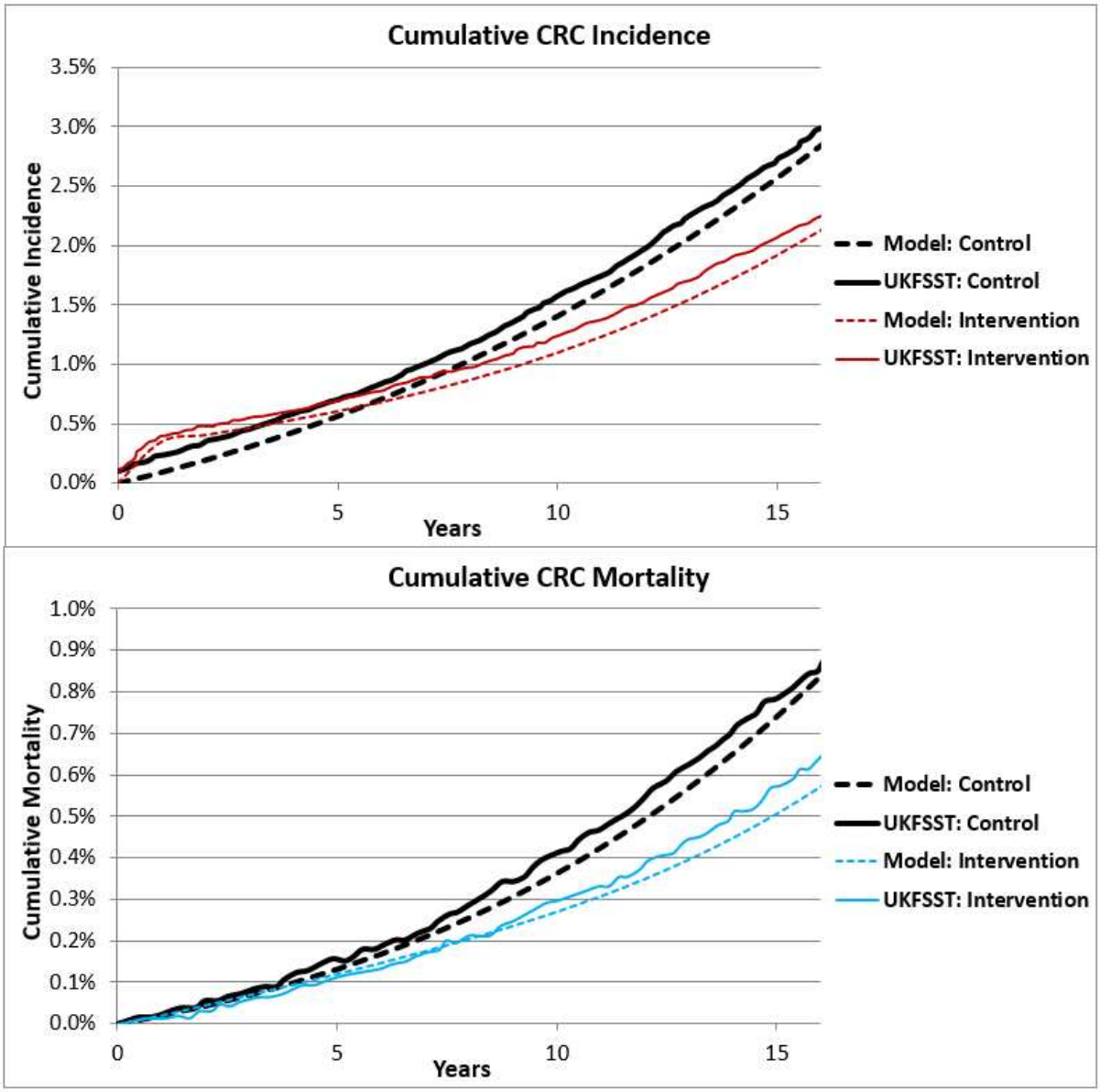


Figure 3: Estimates of cumulative colorectal cancer (CRC) incidence (top) and mortality (bottom) for flexible sigmoidoscopy (FS) screening versus no screening from the UK Flexible Sigmoidoscopy Screening Trial (UKFSST) [9] and the model using current day incidence and mortality data with 17 years of follow-up from baseline screen.



References

1. Public Health England. NHS Bowel Cancer Screening (BCSP) programme. 2015; https://www.gov.uk/topic/population-screening-programmes/bowel. Accessed 27th September, 2016.
2. Whyte S, Chilcott J, Halloran S. Reappraisal of the options for colorectal cancer screening in England. *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland.* 2012;14(9):e547-561.
3. Whyte S, Thomas C, Kearns B, Webster M, Chilcott J. Optimising Bowel Cancer Screening Phase 1: Optimising the cost effectiveness of repeated FIT screening and screening strategies combining bowel scope and FIT screening. *Report to the National Screening Committee* 2017; https://www.sheffield.ac.uk/scharr/sections/heds/discussion-papers/18_04-1.782806.
4. Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Med Decis Making.* 2011;31(4):625-641.
5. Tappenden P, Chilcott JB. Avoiding and identifying errors and other threats to the credibility of health economic models. *Pharmacoeconomics.* 2014;32(10):967-979.
6. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value Health.* 2012;15(6):843-850.
7. Murphy J, Gray A. *The cost-effectiveness of immunochemical faecal occult blood testing vs. guaiac faecal occult blood testing for colorectal cancer screening in the NHS Bowel Cancer Screening Programme.* Health Economics Research Centre, University of Oxford;2015.

8. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut.* 2012;61(7):1036-1040.

9. Atkin W, Wooldrage K, Parkin DM, et al. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. *The Lancet.* 2017;389(10076):1299-1311.

10. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *The Lancet.* 2010;375(9726):1624-1633.

11. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet.* 1996;348(9040):1472-1477.

12. Incidence numbers of colorectal cancer for patients diagnosed between 1996 and 2004 in England, by stage. In: dataset UNC, ed2009.

13. Death Registrations summary tables - England and Wales 2014. 2015; https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathregistrationssummarytablesenglandandwalesreferencetables. Accessed 27th March, 2019.

14. National Life Tables UK 2013-2015. 2016; https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables. Accessed 27th March, 2019.

15. Rachet B. Net survival data for cancer of the Colon and Rectum. In: (ICBP) ICBP, ed2012.

16. Nickerson C. NHS Bowel Cancer Screening Programme (BCSP) Data Extract. In:2016.

17. Downing A, Aravani A, Macleod U, et al. Early mortality from colorectal cancer in England: a retrospective observational study of the factors associated with death in the first year after diagnosis. *Br J Cancer.* 2013;108(3):681-685.

18. McClements PL, Madurasinghe V, Thomson CS, et al. Impact of the UK colorectal cancer screening pilot studies on incidence, stage distribution and mortality trends. *Cancer Epidemiol.* 2012;36(4):e232-242.

19. Rutter CM, Knudsen AB, Marsh TL, et al. Validation of Models Used to Inform Colorectal Cancer Screening Guidelines: Accuracy and Implications. *Med Decis Making.* 2016;36(5):604-614.

20. Croswell JM, Ransohoff DF, Kramer BS. Principles of Cancer Screening: Lessons from History and Study Design Issues. *Semin Oncol.* 2010;37(3):202-215.

21. Kearns B, Whyte S, Chilcott J, Patnick J. Guaiac faecal occult blood test performance at initial and repeat screens in the English Bowel Cancer Screening Programme. *Br J Cancer.* 2014;111(9):1734-1741.

22. Investigators UFSST. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. *Lancet.* 2002;359(9314):1291-1300.

23. Kearns B, Whyte S, Seaman HE, et al. Factors associated with completion of bowel cancer screening and the potential effects of simplifying the screening test algorithm. *Br J Cancer.* 2016;114(3):327-333.

24. Lo SH, Halloran S, Snowball J, Seaman H, Wardle J, von Wagner C. Colorectal cancer screening uptake over three biennial invitation rounds in the English bowel cancer screening programme. *Gut.* 2014;64(2):282-291.

25. von Wagner C, Baio G, Raine R, et al. Inequalities in participation in an organized national colorectal cancer screening programme: results from the first 2.6 million invitations in England. *International journal of epidemiology.* 2011;40(3):712-718.

26. Cancer Research UK. Bowel Cancer Statistics. 2017; http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer. Accessed 27th September 2016.

27. Gavin DR, Valori RM, Anderson JT, Donnelly MT, Williams JG, Swarbrick ET. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. *Gut.* 2013;62(2):242-249.

28. Gill MD, Bramble MG, Rees CJ, Lee TJ, Bradburn DM, Mills SJ. Comparison of screen-detected and interval colorectal cancers in the Bowel Cancer Screening Programme. *Br J Cancer.* 2012;107(3):417-421.

29. Pande R, Froggatt P, Baragwanath P, Harmston C. Survival outcome of patients with screening versus symptomatically detected colorectal cancers. *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland.* 2013;15(1):74-79.

30. Cancer Intervention and Surveillance Modelling Network. 2018; https://cisnet.cancer.gov/. Accessed 27th March, 2019.

31. Frazier AL, Colditz GA, Fuchs CS, Kuntz KM. Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA.* 2000;284(15):1954-1961.

32. Loeve F, Boer GJ, van Oortmarssen M, van Ballegooijen M, Habbema JDF. The MISCAN-COLON Simulation Model for the Evaluation of Colorectal Cancer Screening. *Computers and Biomedical Research.* 1999;32:13-33.

33. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: validation and application. *Cancer Epidemiol Biomarkers Prev.* 2010;19(8):1992-2002.