PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

| TITLE (PROVISIONAL) | Cost-effectiveness of the faecal immunochemical test at a range of positivity thresholds compared with the guaiac faecal occult blood test in the NHS Bowel Cancer Screening Programme in England |
|---------------------|--------------------------------------------------------------------------------------------------------------------------|
| AUTHORS             | Murphy, Jacqueline; Halloran, Stephen; Gray, Alastair                                                                   |

VERSION 1 – REVIEW

| REVIEWER             | Stephanie Coward, PhD Candidate University of Calgary, Canada |
|----------------------|----------------------------------------------------------------|
| REVIEW RETURNED      | 01-May-2017                                                     |

GENERAL COMMENTS

This is a very interesting research project. Below are some points of clarification and changes I would recommend.

Abstract: It would be helpful to have actual numbers within the abstract. I would suggest picking a couple which back up your conclusions.

1. It will help the reader comprehend the paper more easily if the focus of the analysis was the cost-utility analysis and then the remaining costs found were summarized in a Budget Impact Analysis (BIA). It is stated in the title that it is a cost-effectiveness model but the use of QALYs as an outcome denotes a cost-utility model; which is not explicitly stated in the paper.

2. For ease of understanding it would be beneficial if the computer program used to create the model is stated within the manuscript and not in the supplementary material. It is stated where the sensitivity analyses are done in the manuscript but not the main model. Headings could be added to the Uncertainty section that separate PSA from univariate etc. Further,
there appears to be a threshold analysis of cancer management costs and when FIT would no longer be cost saving (in the “Cancer management costs” section in the results) which should be grouped with all the other uncertainty analyses in the methods and results. There should be an uncertainty section in the methods section and then a corresponding one in the results section that contain all pertinent information.

5. What is the benefit of choosing this paper’s SF-6D value versus other possible values? As the QOL measures chosen are not specific to this CRC cohort, was this the best choice there was? What was the assumption that was made from the paper specified, explicitly relating to under which category CRC was classified as it seems that they do not explicitly state cancer related diseases?

6. Costs are only increased to 2013/2014 costs. Since it is now 2017 these numbers should be updated to the most recent numbers (i.e., 2015/2016). Also, there should be a reference to the “Health Services Cost Index.” Further the population estimates are from 2014. Are the 2014 numbers the most recent?

7. The results from the PSA should be contained within the Sensitivity Analysis section instead of the “Cost-effectiveness” section. Also, since you are doing multiple PSAs on different threshold Figure 1 is a very unique graphic that really displays the results well.

8. The rationale and methods in the “Screening test characteristics” section in the results should be in the methods section.

9. It appears in the results that you almost have a complete Budget Impact Analysis but they are scattered through the results section under different headings (i.e., total long term costs). To really be able to elicit change in a health care system there needs to be succinct and “take-home” message that is easily understandable to those involved in policy. To help with this, what I would do, is start with the cost-utility analysis and all associated results, then a section on the sensitivity analyses, and then do a full Budget Impact Analysis with all the costs that you mention in “Screening costs in the first year of screening” and “Long-term colonoscopy resource use” etc.

Overall, this is an interesting paper (and great discussion) but the numbers need to be updated and the information reorganized to allow for a better flow to the paper.

REVIEWER
Rosie Meng
Flinders Centre for Epidemiology and Biostatistics
School of Medicine
Flinders University
Australia

REVIEW RETURNED 18-May-2017

GENERAL COMMENTS
The paper is very well written and the methods are documented thoroughly. Although there are many limitations, authors have conducted a few sensitivity analyses to address the uncertainty in parameters used in models and discussed their impact.

I am curious about how FIT cut-off of 40 μg, 100 μg, 150 μg and 180 μg were chosen in the analysis, and why not other specific cut-off, such as 50 μg, for example. Any particular reason for this please?

It would be good that authors could state the software used for the
analyses, as well as providing an overall detailed model diagram if possible.

**REVIEWER**  
Mary Dallat  
Public Health Agency, Northern Ireland  
**REVIEW RETURNED** 23-May-2017

**GENERAL COMMENTS**  
It's not a particularly 'hard hitting' paper and the methods/model have been used before but nevertheless it's a sound study that clearly shows the additional cost-effectiveness of FIT over gFOBT without losing sight of the service implications.

**REVIEWER**  
Iris Lansdorp-Vogelaar  
Erasmus MC, University Medical Center Rotterdam  
Department of Public Health  
The Netherlands  
**REVIEW RETURNED** 30-May-2017

**GENERAL COMMENTS**  
I want to congratulate the authors on a very clearly written manuscript containing important information for the English bowel cancer screening programme. I have the following suggestions:

1. Results: Suggest to present results per 1,000 or per person in the population. Numbers currently used are not meaningful, because it constitutes a single cohort run and as such absolute numbers do not provide any information but rather distract from the main message.

2. Results: why present disease prevalence, rather than disease incidence? I agree that disease prevalence may be interesting from a cost perspective, but disease incidence is interesting from a burden perspective. I would expect fewer CRC cases with FIT than with gFOBT and suggest to add this outcome.

3. Sensitivity analysis on FIT sensitivity: The model finds FIT to dominate gFOBT despite higher per test costs of FIT versus gFOBT. My explanation for this finding is that this is caused by the higher sensitivity of FIT versus gFOBT, and therefore the higher number of (advanced) CRC prevented and thus more savings from CRC treatment with FIT screening compared to gFOBT. However, in the sensitivity analysis evaluating higher levels of FIT sensitivity suddenly FIT is no longer cost-saving compared to FIT. I don't understand why this is. Please explain.

4. Discussion: In the discussion, the threshold for CRC treatment costs are discussed for which FIT is no longer cost-saving compared to gFOBT. I would stress more how unlikely this situation is, since in the past years costs for CRC treatment have only gone up with the introduction of new systemic therapies, while screening costs have remained relatively stable. Also, on the statement of CRC treatment to vary by age, I would make the concluding remark stronger by adding that even with such cost assumptions FIT screening will probably remain cost-saving.

5. Limitations: Please clarify the statement about screening uptake: does this mean that screening uptake is independent of previous uptake and that therefore every time a random percent of the
population participates in screening: this is a strong assumption which will have huge impact on the effectiveness and cost-effectiveness of screening. This implies that after 8 screening rounds virtually the whole population will have had at least one screening. This way of modeling adherence should be mentioned in the methods section and should be elaborated on in the limitations. Also state how you think this will influence the results of FIT vs gFOBT. Also, I don't understand how you used screening uptake by age as a proxy for screening history.

6. Table 3: Why does the number of kits sent but not returned vary over the different FIT cut-off strategies? Is this random noise?

7. Table 4: You should not report negative values for ICER, because these cannot be interpreted. Also, it should be clarified that the ICER constitute ICERS compared to gFOBT and not to the different levels of FIT cut-offs. Also, the confidence interval for the first row in the INB column is off.

8. Supplementary table 1: Please add sources for the parameters.

9. Supplementary Figure 4: I suggest renaming g_{i}. People without neoplasia cannot be detected with neoplasia.

10. Formulas page 8: a_{i} is substituted by (col_uptake_{i} / col_sens_{i}) / e_{i}. Would make this transformation explicit, so that it is easier to follow. The authors now designate the formula for e_{i}, while in fact a_{i} is replaced and so the above formula is transformed.

11. Calculating relative sensitivity of FIT vs gFOBT: The authors have derived the formulas correctly. However, in essence what they are doing is comparing detection rates with FIT with those of gFOBT correcting for the difference in colonoscopy uptake. This corrected detection rate can also be calculated by multiplying the PPV of the test with its positivity rate. Indeed rearrangement of the final formula gives exactly that. I think this approach is much more intuitive than the complex mathematical approach chosen by the authors and therefore easier to follow.

12. Sensitivity of FIT: As the authors point out, the estimated sensitivity of FIT seems low. However, it is hard to judge, because the authors only present estimates for cut-offs of 40 mg/g feces or higher. I suggest to also add estimates for cut-off of 20 mg/g feces as used in the pilot, because that allows readers to judge the face validity of the model and its assumptions. Also, given the importance of test characteristics to the results, I suggest to include these assumptions in the main manuscript.

13. Calculating relative specificity of FIT vs gFOBT: Again I am unsure why the authors choose this elaborate approach. First, I don't see why probabilities need to be transformed to rates first. Can't you just take the ratio of 1-probability of false-positive test? Given that 1-disease prevalence cancels out of the equation?

14. Quality of life: Ness has reported on utilities for CRC. Ness, Am J Gastro 1999.
Reviewer: 1
Dear Authors,
This is a very interesting research project. Below are some points of clarification and changes I would recommend.

Abstract: It would be helpful to have actual numbers within the abstract. I would suggest picking a couple which back up your conclusions.

> We have amended the abstract text accordingly (see Abstract)

1. It will help the reader comprehend the paper more easily if the focus of the analysis was the cost-utility analysis and then the remaining costs found were summarized in a Budget Impact Analysis (BIA). It is stated in the title that it is a cost-effectiveness model but the use of QALYs as an outcome denotes a cost-utility model; which is not explicitly stated in the paper.

> We have reorganised the Results section into separate Cost-Utility Analysis, Sensitivity Analysis, and Budget Impact Analysis sections to reflect the reviewer’s suggestion.

> We have changed Cost-Effectiveness to Cost-Utility at several points in the manuscript to clarify that this is a CU analysis.

2. For ease of understanding it would be beneficial if the computer program used to create the model is stated within the manuscript and not in the supplementary material. It is stated where the sensitivity analyses are done in the manuscript but not in the supplementary material. It is stated where the sensitivity analyses are done in the manuscript but not the main model.

> We have amended the text accordingly (see section Model Structure)

3. Are there no other types of sensitivity analyses that would aid in understanding the uncertainty contained within the model (i.e., scenario of decreased quality of life associated with increased screening tests or threshold analyses)? It would be interesting to see how the value for money varies with alternate scenarios. For example: “Dis-utilities” (i.e., negative values associated with transitions) could be integrated into the QOLs for tests and procedures in a scenario analysis.

> We have included a range of sensitivity analyses in the Supplementary Information (4: Sensitivity Analysis). We do not think a utility decrement would influence the conclusions due to the short-term nature of procedures compared to the model time horizon. For example, taking a utility decrement of -0.01 over 6 months of CRC treatment equates to -0.005 over one year. Given that not all individuals will require CRC treatment, the resulting difference over a 40 year time horizon will be minimal. We have added some text to the Discussion section (Limitations) about the potential to model utility decrements.

4. It would be better to have all explanations of the uncertainty analyses that were performed in the section entitled “Uncertainty” in methods. Currently the specifics of the PSA are in the “Overview” section and the univariate sensitivity analyses are in “uptake of screening and colonoscopy.” Headings could be added to the Uncertainty section that separate PSA from univariate etc.

Further, there appears to be a threshold analysis of cancer management costs and when FIT would no longer be cost saving (in the “Cancer management costs” section in the results) which should be grouped with all the other uncertainty analyses in the methods and results. There should be an uncertainty section in the methods section and then a corresponding one in the results section that contain all pertinent information.

> We have reorganised the manuscript text accordingly by adding a Sensitivity Analyses subheading to the Results section, between the Cost-Utility Analysis section and the Budget Impact Analysis section. This section contains the PSA and univariate results. There is a corresponding section on Uncertainty in the Methods section to outline which analyses were performed.

5. What is the benefit of choosing this paper’s SF-6D value versus other possible values? As the QOL measures chosen are not specific to this CRC cohort, was this the best choice there was? What was the assumption that was made from the paper specified, explicitly relating to under which category CRC was classified as it seems that they do not explicitly state cancer related diseases?

> We consider this to be the best source given the limited range of values available in the literature as it is also consistent with previous analyses for the NHS Bowel Cancer Screening Programme. We
included the utility weights in the sensitivity analyses (see Supplementary Information – 4: Sensitivity Analyses) and found that uncertainty in these values made very little difference to the results.

6. Costs are only increased to 2013/2014 costs. Since it is now 2017 these numbers should be updated to the most recent numbers (i.e., 2015/2016). Also, there should be a reference to the “Health Services Cost Index.” Further the population estimates are from 2014. Are the 2014 numbers the most recent?

> The unit costs have now been updated to 2015/16 (full unit costs are given in the Supplementary Information), and the most recent population estimated (mid-2015) have been used to re-estimate the results. We have also added a citation for the Health Services Cost Index. The text has been amended accordingly.

7. The results from the PSA should be contained within the Sensitivity Analyses section instead of the “Cost-effectiveness” section. Also, since you are doing multiple PSAs on different threshold Figure 1 is a very unique graphic that really displays the results well.

> We have reorganised the manuscript text accordingly (see section Results). We thank the reviewer for these comments.

8. The rationale and methods in the “Screening test characteristics” section in the results should be in the methods section.

> We have reorganised the manuscript text accordingly (see section Methods: Uncertainty)

9. It appears in the results that you almost have a complete Budget Impact Analysis but they are scattered through the results section under different headings (i.e., total long term costs). To really be able to elicit change in a health care system there needs to be succinct and “take-home” message that is easily understandable to those involved in policy. To help with this, what I would do, is start with the cost-utility analysis and all associated results, then a section on the sensitivity analyses, and then do a full Budget Impact Analysis with all the costs that you mention in “Screening costs in the first year of screening” and “Long-term colonoscopy resource use” etc.

> We have reorganised the manuscript text accordingly as mentioned above (see Results section)

Overall, this is an interesting paper (and great discussion) but the numbers need to be updated and the information reorganized to allow for a better flow to the paper.

> We thank the reviewer for their comments and suggestions.

Reviewer: 2

The paper is very well written and the methods are documented thoroughly. Although there are many limitations, authors have conducted a few sensitivity analyses to address the uncertainty in parameters used in models and discussed their impact.

> We thank the reviewer for their comments.

I am curious about how FIT cut-off of 40 μg, 100 μg, 150 μg and 180 μg were chosen in the analysis, and why not other specific cut-off, such as 50 μg, for example.

Any particular reason for this please?

> We chose the FIT cut-offs to align with the results that were available from the FIT pilot study (Moss, 2016 #42; Moss, 2015 #14). Note we have added an additional threshold (20μg Hb/g faeces) in this revision to align with the pilot results.

It would be good that authors could state the software used for the analyses, as well as providing an overall detailed model diagram if possible.

> We have amended the text to state the software used (see Methods section).

> It was not possible to condense the model diagram into one figure due to its complexity, however we feel that the model diagrams in the Supplementary Information) sufficiently capture the key aspects of the model structure.

Reviewer: 3

It’s not a particularly ‘hard hitting’ paper and the methods/model have been used before but nevertheless it’s a sound study that clearly shows the additional cost-effectiveness of FIT over gFOBT without losing sight of the service implications.

> We thank the reviewer for their comments.

Reviewer: 4
I want to congratulate the authors on a very clearly written manuscript containing important information for the English bowel cancer screening programme. 

> We thank the reviewer for their comments.

I have the following suggestions:

1. Results: Suggest to present results per 1,000 or per person in the population. Numbers currently used are not meaningful, because it constitutes a single cohort run and as such absolute numbers do not provide any information but rather distract from the main message. 

> We feel that it is useful to give the results for the entire cohort for the budget impact analysis, for example to allow for estimation of colonoscopy burden nationally for this cohort. However we have added additional key results to the tables to give the budget impact results per 1000 people invited for screening, and have referred to these in the Results section.

2. Results: why present disease prevalence, rather than disease incidence? I agree that disease prevalence may be interesting from a cost perspective, but disease incidence is interesting from a burden perspective. I would expect fewer CRC cases with FIT than with gFOBT and suggest to add this outcome.

> Due to the Markov structure of the model the incidence rates are not direct outputs of the model, rather the health state distribution gives the number in each state in each cycle. We feel prevalence is an informative measure for illustrating how the burden of disease will affect resource use and costs.

3. Sensitivity analysis on FIT sensitivity: The model finds FIT to dominate gFOBT despite higher per test costs of FIT versus gFOBT. My explanation for this finding is that this is caused by the higher sensitivity of FIT versus gFOBT, and therefore the higher number of (advanced) CRC prevented and thus more savings from CRC treatment with FIT screening compared to gFOBT. However, in the sensitivity analysis evaluating higher levels of FIT sensitivity suddenly FIT is no longer cost-saving compared to FIT. I don't understand why this is. Please explain.

> Although we note that as the sensitivity of FIT increases the colonoscopy costs increase, we have also revised this sensitivity analysis. In the original sensitivity analyses we increased the sensitivity of FIT by 30% (in absolute terms) for LR, HR, and CRC, however we do not feel that this reflects plausible values for LR which is 0.75% in the base case. Using a value of 30.75% in the sensitivity analysis is a very large difference and does not reflect a plausible value. We have replaced this sensitivity analysis with one that only varied the sensitivity for CRC, as the literature suggests sensitivity for CRC could be up to 30% greater than our base case parameters for detecting CRC (but not LR or HR adenomas). The results of the new sensitivity analysis do not have the same result as pointed out by the reviewer that FIT becomes cost saving.

4. Discussion: In the discussion, the threshold for CRC treatment costs are discussed for which FIT is no longer cost-saving compared to gFOBT.

I would stress more how unlikely this situation is, since in the past years costs for CRC treatment have only gone up with the introduction of new systemic therapies, while screening costs have remained relatively stable.

> We have amended the text accordingly (see Discussion)

Also, on the statement of CRC treatment to vary by age, I would make the concluding remark stronger by adding that even with such cost assumptions FIT screening will probably remain cost-saving.

> We have amended the text accordingly (see Discussion)

5. Limitations: Please clarify the statement about screening uptake: does this mean that screening uptake is independent of previous uptake and that therefore every time a random percent of the population participates in screening; this is a strong assumption which will have huge impact on the effectiveness and cost-effectiveness of screening. This implies that after 8 screening rounds virtually the whole population will have had at least one screening. This way of modeling adherence should be mentioned in the methods section and should be elaborated on in the limitations. Also state how you think this will influence the results of FIT vs gFOBT. Also, I don't understand how you used screening uptake by age as a proxy for screening history.

> We have added more information about how uptake is modelled (see Methods and Discussion)

> We have removed the text in the manuscript around using age as a proxy for screening history as
we agree this is unclear, and instead have highlighted that the cohort modelling here may differ from cross-sectional modelling (see Discussion: Limitations).

6. Table 3: Why does the number of kits sent but not returned vary over the different FIT cut-off strategies? Is this random noise?
   > (Now Table 5) Fewer kits are sent over the time horizon for lower FIT thresholds because of differences in factors such as the detection rates (at lower thresholds there are more people in surveillance where they are not being screened every two years), so the cost per person is also slightly lower at lower thresholds.

7. Table 4: You should not report negative values for ICER, because these cannot be interpreted. Also, it should be clarified that the ICER constitute ICERS compared to gFOBT and not to the different levels of FIT cut-offs. Also, the confidence interval for the first row in the INB column is off.
   > We have replaced the negative ICERs with a statement that FIT dominates compared to gFOBT and corrected the confidence interval for the net benefit. We have also amended the column headers to state that the incremental results are compared to gFOBT.

8. Supplementary table 1: Please add sources for the parameters.
   > We have added the sources to Supplementary Table 1

9. Supplementary Figure 4: I suggest renaming g_i. People without neoplasia cannot be detected with neoplasia.
   > Due to changes in the methods to calculate sensitivity and specificity (see below) this section has been removed from the Supplementary Information.

10. Formulas page 8: a_i is substituted by (col_uptake_i*col_sens_i)/e_i. Would make this transformation explicit, so that it is easier to follow. The authors now designate the formula for e_i, while in fact a_i is replaced and so the above formula is transformed.
    > Due to changes in the methods to calculate sensitivity and specificity (see below) this section has been removed from the Supplementary Information.

11. Calculating relative sensitivity of FIT vs gFOBT: The authors have derived the formulas correctly. However, in essence what they are doing is comparing detection rates with FIT with those of gFOBT correcting for the difference in colonoscopy uptake. This corrected detection rate can also be calculated by multiplying the PPV of the test with its positivity rate. Indeed rearrangement of the final formula gives exactly that. I think this approach is much more intuitive than the complex mathematical approach chosen by the authors and therefore easier to follow.
    > We thank the reviewer for these suggestions and have simplified the methods to the ratio of detection rates as suggested. This gives the same numerical values but we agree it is more intuitive.
    (see Supplementary Information: 2. Screening test characteristics)

12. Sensitivity of FIT: As the authors point out, the estimated sensitivity of FIT seems low. However, it is hard to judge, because the authors only present estimates for cut-offs of 40 mg/g feces or higher. I suggest to also add estimates for cut-off of 20 mg/g feces as used in the pilot, because that allows readers to judge the face validity of the model and its assumptions. Also, given the importance of test characteristics to the results, I suggest to include these assumptions in the main manuscript.
    > We have added the results for 20µg Hb/g faeces throughout the manuscript.
    > We have added the assumptions around sensitivity and specificity for the base case analysis to the text of the manuscript (see Methods: Screening Test Characteristics).

13. Calculating relative specificity of FIT vs gFOBT: Again I am unsure why the authors choose this elaborate approach. First, I don't see why probabilities need to be transformed to rates first. Can't you just take the ratio of 1-probability of false-positive test? Given that 1- disease prevalence cancels out of the equation?
    > We thank the reviewer for these suggestions and have simplified the methods for calculating specificity to remove the calculation using rates, and instead using 1- probability of false-positive test (see Supplementary Information: 2. Screening test characteristics).

14. Quality of life: Ness has reported on utilities for CRC. Ness, Am J Gastro 1999.
    > We thank the reviewer for this suggestion. Given the limited range of values available in the
literature we chose to use utility estimates that were consistent with previous analyses for the NHS Bowel Cancer Screening Programme. We included the utility weights in the sensitivity analyses (see Supplementary Information – 4: Sensitivity Analyses) and found that uncertainty in these values made very little difference to the results.

VERSION 2 – REVIEW

**REVIEWER**
Xingqiong Meng  
Research Fellow  
College of Medicine and Public Health,  
Flinders University  
I have no conflict of interest.

**REVIEW RETURNED**
21-Jul-2017

**GENERAL COMMENTS**
The authors are responsive to the comments and the manuscript is improved.

**REVIEWER**
Iris Lansdorp-Vogelaar  
Erasmus MC University Medical Center Rotterdam  
The Netherlands

**REVIEW RETURNED**
31-Jul-2017

**GENERAL COMMENTS**
1. The authors have nicely separated out the cost-utility analysis from the budget impact analysis. However, I feel that the budget impact analysis, is not a true budget impact analysis, since it concerns costs and impact of only one cohort (i.e. 60-yr olds followed over lifetime) rather than the screened population as a whole (also including 62, 64,..., 74 year olds. I realize that the authors have built a cohort model, however by combining the results at different ages, the authors can still determine the results in the first year(s) of the program at a population level. Ladabaum et al have done something similar in their publication in Clin Gastroenterol Hepatol. in 2004.

2. In UK practice FIT screening starting at age 60 years will come after a once-only sigmoidoscopy at age 55 years. In this analysis, the authors did not model this first sigmoidoscopy. However, the results of FIT and/or gFOBT will depend substantially on whether sigmoidoscopy has been performed prior. If possible, the authors should consider modeling the sigmoidoscopy at age 55, or at least discuss the limitation of this omission in the discussion.

3. As requested by me, the authors have added the reference for the source data of the model to Supplementary Table 1. My intention with this suggestion was to have references to original data on which parameters were based. I now realize that all parameters were directly taken from Whyte et al. As such, adding the source in a separate column does not add to the table. I suggest adding the source as a footnote to the table instead.

4. Suggest adding a supplementary table comparing observed and predicted model detection rates to allow assessment of face validity of the model.
Reviewer: 2
Reviewer Name: Xingqiong Meng
Institution and Country: Research Fellow, College of Medicine and Public Health, Flinders University
Please state any competing interests: I have no conflict of interest.
Please leave your comments for the authors below
The authors are responsive to the comments and the manuscript is improved.

[Response: We thank the reviewer for these comments.]

Reviewer: 4
Reviewer Name: Iris Lansdorp-Vogelaar
Institution and Country: Erasmus MC University Medical Center Rotterdam, The Netherlands
Please state any competing interests: None declared
Please leave your comments for the authors below
The authors have done a great job addressing the reviewers’ concerns.

[Response: We thank the reviewer for these comments.]

I have few remaining suggestions:
1. The authors have nicely separated out the cost-utility analysis from the budget impact analysis. However, I feel that the budget impact analysis, is not a true budget impact analysis, since it concerns costs and impact of only one cohort (i.e. 60-yr olds followed over lifetime) rather than the screened population as a whole (also including 62, 64,..., 74 year olds. I realize that the authors have built a cohort model, however by combining the results at different ages, the authors can still determine the results in the first year(s) of the program at a population level. Ladabaum et al have done something similar in their publication in Clin Gastroenterol Hepatol. in 2004.

[Response: We have added an estimate for the annual cost in a steady state, similar to the analysis by Ladabaum as noted by the reviewer. See Methods and Discussion.
We have also moved the text explaining the methods for the budget impact analysis from the Results section to the Methods section.
We have also adjusted the original budget impact results (highlighted in Tables 2 and 3) and the accompanying text to make a correction to the assumed population size at age 60 based on data from the Office for National Statistics.]

2. In UK practice FIT screening starting at age 60 years will come after a once-only sigmoidoscopy at age 55 years. In this analysis, the authors did not model this first sigmoidoscopy. However, the results of FIT and/or gFOBT will depend substantially on whether sigmoidoscopy has been performed prior. If possible, the authors should consider modelling the sigmoidoscopy at age 55, or at least discuss the limitation of this omission in the discussion.

[Response: We hope the reviewer accepts that to model the costs and effects of the once-only sigmoidoscopy at age 55 would be a major study in its own right and would go well beyond the scope of this study. Instead we have adopted the reviewer’s second suggestion, and extended in the Discussion section our existing mention of flexible sigmoidoscopy screening and its possible impact on the results of the present analysis.]

3. As requested by me, the authors have added the reference for the source data of the model to Supplementary Table 1. My intention with this suggestion was to have references to original data on which parameters were based. I now realize that all parameters were directly taken from Whyte et al. As such, adding the source in a separate column does not add to the table. I suggest adding the
source as a footnote to the table instead.

[Response: We have amended Table 1 in the supplementary information and added a footnote.]

4. Suggest adding a supplementary table comparing observed and predicted model detection rates to allow assessment of face validity of the model.

[Response: We have added tables in the Supplementary Information (Supplementary Tables 14 and 15) comparing the model results to observational data from the bowel cancer screening programme on the results of screening episodes (CRC and adenoma detection rates). We have also added text to the Discussion section.]

**VERSION 3 – REVIEW**

| REVIEWER       | Iris Lansdorp-Vogelaar  
|                | Erasmus MC, The Netherlands |
| REVIEW RETURNED | 22-Aug-2017 |

**GENERAL COMMENTS**
The authors have adequately addressed my remaining concerns.