## 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)** | What are the cost-savings and health benefits of improving detection and management for six high cardiovascular risk conditions in England? An economic evaluation |
| **AUTHORS** | Thomas, Chloe; Brennan, Alan; Goka, Edward; Squires, Hazel; Brenner, Gilly; Bagguley, David; Buckley Woods, Helen; Gillett, Michael; Leaviss, J; Clowes, Mark; Heathcote, Laura; Cooper, Katy; Breeze, Penny |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Dario Gregori<br>University of Padova, Italy |
| **REVIEW RETURNED** | 05-Mar-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | This a very well-written paper addressing an important topic. The overall information and the level of detail provided to the reader are absolutely outstanding.<br>The work, as recognized by the authors, suffer from the limited capability of translating its finding into policy indications, also in terms of budgeting dedicated to the prevention and health care. This may particularly impact short-term policy decisions when most of the intervention models are likely to be not effective. I belive that discussion should be widened in this respect.<br>As minor remarks, i couldn't find specific detail on how correlation among chronic diseases considered was managed. Since most of them are likely to occur simultaneously on the same patient, in particular with aging, this may impact overall diagnostic and management trajectories.<br>The models presented and discussed are complex to follow in detail from a statistical perspective. I would recommend to include the R-code used in the analyses as supplementary material. this could help also in reproducing analyses in other similar contexts. |

| | |
|---|---|
| **REVIEWER** | Jan Sorensen<br>Healthcare Outcome Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland |
| **REVIEW RETURNED** | 02-May-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is interesting and important research that is highly relevant for the debate about prioritising prevention in healthcare. There is an impressive amount of work behind this report and it is clearly conducted by a large group of highly competent multidisciplinary researchers.<br>It is clearly written and includes all relevant information to understand the main results with the detailed analysis and methodological and empirical choices available in the substantive |

| | |
|---|---|
| | supplementary material.<br><br>The main choices are summarised including the time perspective 25 years clearly explained. One thing I did not find entirely clear is on p15 l 60-69 where it talks about increased costs related to ESRF, OA, Breast and Bowel cancer, depression, dementia and major bleed. It seems as the development of these diseases are included in the modelling and the cost, and it is reported that they results in "small net increases". I find that surprising when the CVD to a large extent is controlled and survival time increased, that the increases in prevalence and costs of these diseases are small. Some are chronic and may last for remaining life time and associated with substantial costs. I suspect that this part of the model is less detailed than the other CVD parts, and that there would be used less detailed empirical data and complicated algorithms. It could be useful to dilute to the size (more than small) of these effects and perhaps discuss the strength of this part of the modelling in the discussion.<br><br>In the tables I am surprised of the wide credible intervals.The Net total costs all include 0. It is good to provide an assessment of the uncertainty in the results, but if the credible intervals are interpreted as normally, then the importance of the findings disappears. I am not sure what to suggest as I suspect this is a reflection of the specified uncertainty in the underlying assumptions. I am really puzzled about the reason and implications.<br><br>Related to the above, I am wondering about the stability of the model if the seed and number of run (2000) are changed. Hopefully, there will be no substantial change, but it might be worth checking. The choice of 2000 runs is different from what I would have expected (1000 or 10000) but that has probably no serious implication, although a higher number of runs may reduce the credible interval of the simulations.<br><br>A couple of minor observations:<br>The sequence of authors is inconsistent including the contribution section.<br><br>On p 20 li 5 "galvanise improvement" is a funny term - perhaps a reminisicent of the steel industry in Sheffield :-) |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer: 1
Reviewer Name: Dario Gregori
Institution and Country: University of Padova, Italy
Please state any competing interests or state ‘None declared’: None declared

This a very well-written paper addressing an important topic. The overall information and the level of detail provided to the reader are absolutely outstanding.

*We thank the reviewer for the very positive comments.*

The work, as recognized by the authors, suffer from the limited capability of translating its finding into policy indications, also in terms of budgeting dedicated to the prevention and health care. This may

particularly impact short-term policy decisions when most of the intervention models are likely to be not effective. I believe that discussion should be widened in this respect.

*We agree with the reviewer that this is an important point to make and have added the following sentence to the discussion paragraph three:*

***'A further barrier to improvement is the short-term nature of much decision making, when improvements to health and cost-savings will not be realised until far into the future.'***

As minor remarks, I couldn't find specific detail on how correlation among chronic diseases considered was managed. Since most of them are likely to occur simultaneously on the same patient, in particular with aging, this may impact overall diagnostic and management trajectories.

*There are several ways in which correlation between diseases is considered in the model – this is described in full in the supplemental methods but we agree that a summary of it would be helpful in the methods of the main manuscript and have now added an abbreviated summary of the following explanation at the end of the Model Development and Assumptions section.*

*Firstly, most of the modelled conditions are represented in the baseline population which is based on a real and representative population for England, so there is correlation between conditions from the start of the model, although this is better for some conditions than others (for example hypertension and diabetes are well defined through metabolic measurements, but atrial fibrillation is not so patients starting with this at baseline are chosen randomly from those who have the highest scores using the Framingham atrial fibrillation risk equation). Secondly, the trajectories of metabolic risk factors (BMI, blood pressure, cholesterol and HbA1c) are correlated with each other and with other patient characteristics based on statistical analysis of longitudinal data. This means that patient risk for one condition (such as hypertension) is correlated with risk for another (such as high cholesterol or diabetes). Thirdly, many of the risk equations used in the model to determine risk of disease for one condition have coefficients relating to one or more of the other conditions. For example risk of chronic kidney disease is higher in people with hypertension, high cholesterol and diabetes. Fourthly, some diagnostic criteria for a condition vary depending upon whether or not another underlying condition is present. For example, the criteria for diagnosing hypertension in people with diabetes is a slightly lower blood pressure than in people without diabetes and once people are diagnosed with diabetes, guidelines state that they should have blood pressure checked annually, so are more likely to be diagnosed with hypertension if they have it.*

*One area in particular where correlations were not included due to lack of evidence is in actual treatment effects for management interventions. We did not find any good evidence about differential treatment effects in people with additional comorbid conditions or undergoing multiple treatments (for example whether anti-hypertensives work more or less well in people with diabetes or people also being treated with statins). This is a limitation of the model.*

The models presented and discussed are complex to follow in detail from a statistical perspective. I would recommend to include the R-code used in the analyses as supplementary material. This could help also in reproducing analyses in other similar contexts.

*We agree that including a link to R-code (perhaps in GitHub rather than as supplementary material) would be extremely useful and we are in the process of working out how we can do this for our models in general in the future, whilst retaining intellectual property rights. However, a considerable amount of work would be required to make it understandable and usable without error, as the code is very complex (e.g. providing more commenting throughout the code, making code into packages and full code documentation). At the moment therefore our models are not open access and so we cannot*

*provide the code as supplementary material.*

Reviewer: 2
Reviewer Name: Jan Sorensen
Institution and Country: Healthcare Outcome Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland
Please state any competing interests or state 'None declared': None to declare

This is interesting and important research that is highly relevant for the debate about prioritising prevention in healthcare. There is an impressive amount of work behind this report and it is clearly conducted by a large group of highly competent multidisciplinary researchers.
It is clearly written and includes all relevant information to understand the main results with the detailed analysis and methodological and empirical choices available in the substantive supplementary material.

*We thank the reviewer for the very positive comments.*

The main choices are summarised including the time perspective 25 years clearly explained. One thing I did not find entirely clear is on p15 l 60-69 where it talks about increased costs related to ESRF, OA, Breast and Bowel cancer, depression, dementia and major bleed. It seems as the development of these diseases are included in the modelling and the cost, and it is reported that they results in "small net increases". I find that surprising when the CVD to a large extent is controlled and survival time increased, that the increases in prevalence and costs of these diseases are small. Some are chronic and may last for remaining life time and associated with substantial costs. I suspect that this part of the model is less detailed than the other CVD parts, and that there would be used less detailed empirical data and complicated algorithms. It could be useful to dilute to the size (more than small) of these effects and perhaps discuss the strength of this part of the modelling in the discussion.

*Firstly, the actual value of the costs is available in supplementary Table 1 and a reference to that table has now been added to direct the reader to the actual values if they are interested. Note that the costs are small in the long run relative to the savings from CVD rather than small in absolute terms, so this has also been clarified at this point in the manuscript.*

*Secondly, the reviewer is correct in thinking that these other conditions are not modelled in quite the detail of CVD and it is possible that this might be part of the reason for such costs being relatively low. Another reason may be that there are many other conditions not included in the modelling (only conditions with some relationship to BMI or diabetes were included), which would be also be expected to increase in prevalence (and hence cost) with greater survival from CVD. Not including all of these conditions is a limitation of the model. A sentence describing this has now been added to the discussion in paragraph two as follows:*

*'An assessment of some of the competing risks for mortality and morbidity is also possible due to the inclusion in the model of a range of additional conditions including dementia, depression, some cancers and osteoarthritis;* ***although it is important to note that these conditions have been modelled in a less sophisticated way than cardiovascular disease, and that the model does not include all competing risks, so there may be additional long-term costs to preventing CVD not included in the model'****.*

In the tables I am surprised of the wide credible intervals. The Net total costs all include 0. It is good to provide an assessment of the uncertainty in the results, but if the credible intervals are interpreted as normally, then the importance of the findings disappears. I am not sure what to suggest as I

suspect this is a reflection of the specified uncertainty in the underlying assumptions. I am really puzzled about the reason and implications.

*The wide credible intervals are a consequence of the large number of uncertain parameters in the model (several hundred) and the inability to accurately correlate them all with each other (some have known correlations that have been included, but most are considered to be independent).*

*Interpretation of credible intervals in health economic modelling is not usually treated in the same way as interpretation of confidence intervals in statistical analysis, whereby only results which are significant at the 95% level are assumed to be significant. Instead, health economic modelling is used to inform decisions where there are two or more options, with each option having a probability of being the most cost-effective or cost-saving (in total adding up to 100%).*

*This means that whilst all the net total costs include 0, they all have a higher probability of either being less than 0 (cost saving) or greater than 0 (costly), compared to current care, which can be seen from the spread of the 95% CI. Even if the probability that an intervention is cost-saving is only 60%, compared to not intervening, the probability that not intervening is cost-saving compared to intervening will only be 40%, which swings the balance of the decision towards intervening. This is a high level of uncertainty but it can still help decisions to be made, or alternatively, more research to be done to reduce the uncertainty before making a decision.*

*The following sentence has been added to the discussion to explain this briefly:*

***'Note that most individual results are uncertain due to the large number of uncertain parameters in the model and the lack of information about how these are correlated; however in most cases the balance of probabilities is in favour of additional detection and management over the long term'***

Related to the above, I am wondering about the stability of the model if the seed and number of run (2000) are changed. Hopefully, there will be no substantial change, but it might be worth checking. The choice of 2000 runs is different from what I would have expected (1000 or 10000) but that has probably no serious implication, although a higher number of runs may reduce the credible interval of the simulations.

*Increasing the number of runs will reduce stochastic (random) uncertainty, but it will not have much impact on the credible intervals which reflect parameter uncertainty. This is illustrated in the following table which shows the mean and 95% credible intervals for 5 year total costs for the 100% detection scenario (the value in the first row and column of Table 2 in the manuscript), for different numbers of PSA runs up to 2000. This particular result appears to be fairly stable with even as few as 200 PSA runs. There are hundreds of results presented in the paper, so it isn't possible to check this for all of them and those based on smaller numbers of people (e.g. subgroup results) will be less stable due to higher stochastic uncertainty; however 2000 runs was chosen as something that should be sufficiently stable for the vast majority of results without being overly time consuming to run.*

**Example Table illustrating stability of manuscript Table 2 line one column one result (5 year Total costs for the 100% detection scenario) at different PSA runs.**

| *Number PSA Runs* | *Mean* | *Lower 95% CI* | *Upper 95% CI* |
|---|---|---|---|
| *50* | *£2.14b* | *-£3.03b* | *£5.01b* |
| *100* | *£2.34b* | *-£2.15b* | *£5.99b* |
| *200* | *£2.42b* | *-£3.48b* | *£6.07b* |
| *400* | *£2.39b* | *-£3.64b* | *£5.93b* |
| *800* | *£2.39b* | *-£3.39b* | *£5.90b* |
| *1500* | *£2.47b* | *-£3.21b* | *£5.94b* |
| *2000* | *£2.43b* | *-£3.36b* | *£5.98b* |

A couple of minor observations:
The sequence of authors is inconsistent including the contribution section.

*These have now been changed to reflect the author sequence.*

On p 20 li 5 "galvanise improvement" is a funny term - perhaps a reminiscent of the steel industry in Sheffield :-)

*Thanks for making me laugh! There are indeed two meanings of galvanise, with one relating to metal working and the other meaning 'to shock or excite someone into taking action'. The meaning here is the second definition, which I think it appropriate, so have left it as it is.*