Author’s response to reviews

Title: The impact generated by public and charity funded research in the UK: A systematic literature review

Authors:
Daniela Gomes (d.gomes@nhs.net)
Charitini Stavropoulou (C.Stavropoulou@city.ac.uk)

Version: 1 Date: 29 Nov 2018

Author’s response to reviews:

Reviewer #1: HRPS-D-18-00204: The impact generated by public and charity funded research in the UK: A systematic literature review Reviewer's comments

This review is highly relevant for Health Research Policy and Systems, and is generally well-written. It has potential for publication as it has identified a range of papers on the impact generated by public and charity funded health research in the UK published between 2006 and 2017, and made progress in collating and analysing the findings from them. However, in order to contribute a useful critical analysis I believe a rather more precise, detailed and nuanced analysis is required that builds on some of the key points already included.

We would like to thank the reviewer for such a thorough, constructive and detailed feedback. We welcome all the comments, which we address point-by-point below. We hope the reviewer finds merit in our revision.

In terms of precision, in several of the numbered points below I indicate how the accuracy of the paper might be improved. For example, by addressing the fact that two papers describing exactly the same study findings have been included, and, further, are given different scores from each other for the risk of both funding bias and selection bias. Additionally, it appears as if at least one included paper does not purport to be an assessment of impacts, but rather is an interesting analysis of research inputs. Some of the definitions of categories of risk look questionable, for example the one in relation to funding. This definition results in Sainty (2013) being the only paper categorised as low risk of bias for funding, and yet the author conducted the impact assessment as the paid R&D manager of the research programme whose impact she was assessing.

The reviewer is right in that we needed to reconsider whether both papers describing the findings of the same study should be included in our analysis. For reasons explained in more detail in our response to comment 4 below, we decided to include only one. The reason they were given different scores in the original submission has to do with the lack of clarity of the previous risk tool, not with the accuracy of our analysis. We have now refined the tool in a way that such discrepancies would not have occurred had we decided to include both studies.
Regarding the Hall et al (2012) study, we believe it fits our criteria for inclusion and decided to keep it in the analysis (please see our detailed response to comment 7 below).

Finally, the reviewer is absolutely right regarding Sainty (2013). As we explained in our response to comment 5 below, we based our rating on the funding declaration in the acknowledgments of the paper, not the conflict of interest of the author. As a result, we have changed the risk of bias tool to reflect funding given to the authors as well as funding given to the study, and we have rated the Sainty (2013) study as one of high risk of funding bias.

The current review also makes a harsh assessment that most of the papers have a 'high risk' of reporting bias and 'failed to provide supporting evidence when making claims'. In various comments below related to specific points in the text I highlight how this current review itself appears to have overlooked data that are provided in some of the studies. Such omissions, in turn, might raise some doubts about the accuracy of some of the scores given for the risk of reporting bias. The authors might feel, as they imply in the limitations in relation to the article by Morgan Jones et al, that they did not have the resources to tackle the challenge of extracting precise data from each of the studies. However, in that case, it is unclear how the authors of the current review could conclude that little evidence had been provided, and that, therefore, such papers should be classified as 'high risk' for reporting bias. My comments at various stages below should help the authors identify some of the missing, and/or inaccurate, data in their analysis.

We now clarify more in depth what we mean by “reporting bias” and provide more evidence in the text to justify how we reached the assessment of the studies. As we explain with more details later on, and describe in Table 1, reporting bias is not about how long a report is. It is about a) whether all aspects of the study are clearly described and, more importantly, b) whether the reported data are consistently more favourable of the outcome/funding in the analysis. Our analysis showed that the later bias was present in quite a few of the studies we reviewed.

Having said that, we absolutely agree that some studies have put herculean effort in presenting their results, and we highlight this in our discussion.

Furthermore, to provide a more nuanced analysis it would be useful to consider issues such as the balance between the desire to provide a detailed analysis, as in some of the included papers, and the cost of doing so. The latter concern helped inform some of the new approaches to impact assessment.

We absolutely agree with the reviewer on this point. Below we clarify how we address it in our paper.

Furthermore, the review could usefully build on important points made at the start of the final paragraph of the Conclusion about research activity having a degree of risk and unpredictability, and the consequent implications for impact assessment. It might be helpful if these points were brought into the Background section and informed the analysis in various ways. For example, while the aims of each study are described in the Table, the implications of some of them do not seem to be fully taken into account in various parts of the analysis.
This is a very valid point and we have now included in the introduction a paragraph around the uncertainty of doing research, the challenges of achieving its aims and the implications of this in assessing the impact of research. Please see page 3 of the revised manuscript.

If research activities are unpredictable, and if, as with the Morgan Jones et al (2016) study, the stated aim is to identify 'evidence of NIHR funded research which has generated benefits to and wider impacts on health', then it could be argued that sometimes a selective selection approach is the most appropriate one to use to achieve the aim.

The case of Morgan Jones et al (2016) is a good example to demonstrate the point of selection bias as presented in our analysis. Indeed, the “study was commissioned as a synthesis of impacts and benefits, not an evaluation” (p.4) to celebrate the successes of the NIHR at 10 years. In that sense, the authors were right in following a selective selection approach and were very transparent in stating this as a limitation in the report. But of course, the selection of studies was not representative of the projects funded by the NIHR, and this is the point we are trying to make. Six more studies followed a similar approach.

Comments about the need for a more nuanced approach are also incorporated in the list of numbered points below:

1. P.2 Abstract: might need to be revised following various revisions to the text.

The abstract has changed to reflect the changes in the text below, mainly reflecting on risk of bias assessment. Please see page 2 of the revised manuscript.

2. P.3: Background: Various points about the nature of research, and the implications for impact assessment could usefully be brought in here, along with a recognition that the purposes of research impact assessments might vary and different categories of research are more amenable to impact assessment than others. It might also be useful to set the paper in the context of the existing literature to a greater extent, for example by drawing on the analysis in the reviews of the literature in the Hanney et al (2007) paper and Raftery et al (2016) - both of which are already referenced in the current review.

As mentioned above, this is a valid point and we have enriched the background to highlight this. Please see page 3 of the revised manuscript.

3. P.4, lines 30-35: it would be helpful to explain why the journals hand-searched were thought to be 'key' when they did not provide any of the papers included in the review, and only BMC Health Services Research provided any papers referenced at all in the study.
Indeed, the choice of the journals could have been broader and we now mention this as a limitation of the study in our discussion. Please see page 21 of the revised manuscript.

4. **P4, Study selection:** it would be helpful to explain the approach to duplicate publications from the same study. Hanney et al (2007) is a Health Technology Assessment report and such reports have a highly unusual, but significant role. They provide the opportunity for research teams to present extremely detailed accounts of the aims, methods, findings and implications of their studies. In many ways they are 'monographs', but because they undergo rigorous peer-review they have been accepted as articles in the journal Health Technology Assessment which was given an ISI Web of Knowledge Journal Impact Factor. However, because these are such detailed publications, the HTA Programme encourages authors also to publish much briefer versions of their papers in traditional journals. Therefore, in this instance, Raftery et al (2009) is a brief journal article on the same data as described in Hanney et al (2007).

We are grateful to the reviewer for bringing this up, as it was a point that we discussed extensively. Our approach was to remove duplicate publications not studies. But we agree that ultimately, the two papers were two different outcomes of the same study, and this becomes evident in our results, when Raftery et al (2009) is mentioned mainly to confirm the findings of Hanney et al (2007). We have therefore decided to remove Raftery et al (2009) as a duplicate study and keep Hanney et al (2007) as the original and more detailed study. We have added a clear statement in the results section to clarify this choice. Please see page 7 of the revised manuscript.

5. **P6, Table 1:** Further refinement is required of the categorisations for funding bias. The low risk category is defined as 'The study did not receive any funding'. However, the only study analysed as being in this category (Sainty et al) raises considerable doubts about the adequacy of this definition as it stands. Sainty herself points out the conflicts of interest in her study, because she conducted it as part of her job as the R&D manager of the research fund whose projects she was assessing. So, it seems odd that those circumstances should lead to a 'low risk' categorisation, but studies that were conducted totally, or primarily, by research teams independent of the sponsor (albeit funded by the sponsor) should be categorised as being 'high risk', (including Bunn et al; Guthrie et al, 2015, and 2016; Hanney et al, 2017 and 2013; Lichten et al, 2017; Morgan Jones et al, 2016). Neither is it explained why a study receiving funding from a different funding body than the one it is evaluating should be categorised as 'medium risk' rather than 'low risk'. Similarly, is it appropriate that a study where the funding is not stated is viewed as being only 'medium risk', whereas studies that are transparent about their funding from the sponsor are 'high risk'? This also seems to contradict the approach taken on selection bias, where if the response rate is not reported, the study is viewed as being 'high risk'.

- We have now refined the ‘funding bias’ categories and we would like to thank the reviewer for giving us constructive suggestions to do so. High risk is attached to a study funded/commissioned exclusively by the funding body it evaluates; medium...
risk if funded partly by it and low risk if no funding has been given by this body. This categorisation clarifies that funding bias relates to the potential conflicts, real or perceived, that may occur when the body reviewed is fully, or partly financing the study or its authors. The categorisation now leaves aside funding provided by other bodies.

- The reviewer is absolutely right regarding Sainty et al (2013). Our assessment of the funding bias was based on the author’s acknowledgment on funding not on conflict of interest. We have now changed our assessment accordingly. More importantly, as stated above, we have changed the ‘funding bias’ categories to reflect that high risk can occur not only if the funding body funds the study, but also its authors.

- For the study that did not report funding sources, we put “N/A” in this category, but ranked it as high risk of reporting bias for not declaring important information.

6. P.6, Study inclusion: compared to the review in Hanney et al (2007) this current study has missing data because it does not list all the papers (48) taken to full review. Inclusion of such a list would allow the informed reader to check, for example, whether Kuruvilla et al (2007) had been identified in the search, but excluded, or had not been identified (this is the paper that describes the application of the approach presented in Kuruvilla et al, 2006).

Both Kuruvilla et al (2006) and Kuruvilla et al (2007) were identified through database searching. However, our first criterion for inclusion of a study was that it “analysed explicitly the impact generated by research funding provided by public and charity bodies in the UK”. Kuruvilla et al (2007) analysed projects whose funders were not necessarily from the UK, but included the European Union, the United States Agency for International Development among others. The study was therefore excluded. We would be happy to share both the list of 62 studies, which were assessed looking at abstract and title and the 48 studies taken to full review if the editor agrees.

7. P6/7, Study characteristics: As noted above, I do not believe the study by Hall et al is a research impact study, and neither do I believe it is accurate to state that it looks at ‘the effect of health research on tobacco and disease burden’. What it does do is make an important analysis of how far the funding for tobacco research matches the burden of disease associated with tobacco. Therefore, it is an important analysis of research inputs, not impact. As also noted, it might be useful to discuss the purpose of the assessment studies because of the implications that might have for the appropriateness of different methods. Furthermore, I think the statement that the design of the studies was such that none of the studies could establish causality between research funding and various outcomes might need some disaggregation. Will there not be differences in terms of
clarity of causality, for example, between, on the one hand research funding and the production of articles, and on the other hand research funding and improved health?

The Hall et al (2016) study’s aim is “to assess the relation between investment in UK health research and disease burden, with a particular focus on tobacco research and burden of tobacco-related disease”. All funding bodies they explored were government and charitable organisations in the UK. It therefore, meets the inclusion criteria for the review and was included. We would agree with the reviewer though, that by no means their analysis can establish causality, but as it is mentioned in the text, this is a wider issue with the literature.

P.8/9: Risk of bias and Table 2: it is good that the authors picked up that the Hanney et al (2007) paper had used a random selection approach, and thus was low risk for selection bias, but it is not clear why for the identical study (Raftery et al) was categorised as being medium risk. In addition to the points made earlier, much further analysis is required around the statement that 'many studies stated that impact was generated but provided little supporting evidence (9/15 or 60%).' Presumably this refers to the 9 studies put in the high risk category for reporting bias. A proper analysis of the Guthrie et al (2015) paper, for example, would reveal that this was published as a report in the Health Technology Assessment series. It is over 300 pages long and contains an enormous amount of supporting evidence, as noted in the following extract from their paper (Guthrie et al, 2015, p.xxi):

'We explored a wide range of impacts resulting from HTA programme-funded research and the HTA programme. We carried out an analysis of impact across the HTA programme using the following methods:

* Interviews (n=20) Senior stakeholders from academia, policy-making organisations and the HTA programme.

* Bibliometric analysis Citation analysis of publications (n=1087) arising from HTA programme-funded research.

* Researchfish survey Electronic survey of all HTA grant holders (n=619) [excluding Technology Assessment Reports (TARs)].

* Payback case studies (n=12) In-depth case studies of HTA programme-funded research, which included document review, interviews and bibliometric analysis.

This multi-method study allowed us synthesise data from multiple sources to identify key findings regarding the impact of the HTA programme.'

It should be noted that these data include an analysis of all the data provided by the 619 relevant projects in their submission to Researchfish (ie, a method which is promoted in the Conclusion of the current review as a way to provide a more systematic data collection). But, the Guthrie et al study goes much further because they have also supplemented the Researchfish data by adding
a major citation analysis exercise and 12 extremely detailed case studies. I find it difficult to see how it could be correct to conclude that the Guthrie et al study 'provided little supporting evidence'. Similarly, the HTA report by Hanney et al (2007) is 200 pages long and contains a full analysis of the evidence from a comprehensive survey and 16 detailed case studies, each of which uses the elements of the Payback Framework to report the evidence from an interview, extensive documentary analysis and triangulation. Furthermore, various other papers report considerable evidence from extensive, and often diverse, data collection activities. These studies include: Hanney et al, 2013; Lichten et al, 2017; Morgan Jones et al, 2016; Peckham et al, 2008.

As mentioned earlier, we clarify that reporting bias is not about how lengthy the report may be, although in some cases the very brief description of say the methods section (McCrae et al 2012), meant very little was known about important aspects of the study. Reporting bias, as stated in Table 1, considers whether a) all aspects of the study are clearly described and b) whether the reported data are consistently more favourable of the outcome/funding in the analysis. Our analysis showed that, even very lengthy reports tended to highlight more the positive evidence on impact and less the lack of it.

8. P.9, line 46: after stating the average number of publications reported by Hanney et al (2007) for the HTA projects, the current paper goes on to state: 'Raftery and colleagues [15] confirmed the results and, in addition, showed the mean number of conference presentations was 5.2 per project.' As noted, there is a question as to whether Raftery et al, as a duplicate study, should be included at all, but, irrespective of that, this text seems to imply that Raftery et al provided additional information in relation to the number of conference papers that was not in Hanney et al. In fact, the opposite was the case with Hanney et al not only providing the average 5.2 per project figure, but also the figure for each of the 3 types of projects included in the analysis that came to an overall average of 5.2.

As mentioned earlier, the study by Raftery et al (2009) has now been removed from the analysis.

9. P.11, line 55: the final paragraph refers to 'five' of the reviewed studies providing a series of examples through which funded research has contributed to the development of research capacity. It misses out some further examples. For example, while it was not the most important payback category in the Hanney et al (2007) study, p.55 provides the total number of qualifications gained according to the survey, and each of the 16 case studies has capacity building etc as a heading, and reports whether or not, and how, it arose in the particular case.

Thank you for pointing this out. We have added the study in the analysis. Please see page 12 of the revised manuscript.
10. P.14, top line: it is misleading to suggest just 2 studies used a survey to identify impacts on informing policy. Additionally, at least 2 more studies drew on survey evidence as part of a mixed-method approach to assessing the impact on policy, namely Guthrie et al (2015), who used the Researchfish survey data, and Hanney et al (2007).

Thank you for bringing this to our attention. We have now amended the text and included the two studies on page 14 of the revised manuscript.

11. P.15, top paragraph: the text correctly notes the important role for research in informing policy that was recorded in the assessments by Guthrie et al and Hanney et al. The text also correctly notes some of types of policy informed by HTA research. In order fully to correct the point made earlier in the text by the authors (and repeated later) that limited supporting evidence was provided, it would be worth the authors of the current text going further. They could appropriately note that the case study assessments on the impact of the HTA programme went into considerable detail about the exact names of the policy documents informed by specific HTA projects, and the precise issues in the documents that were influenced by the specific research.

The reviewer is right, and we have added a sentence at the end of that paragraph to note that indeed the study provided very detailed reference to documents and names. Please see page 15 of the revised document.

12. P.15, middle para, last line: I don’t think the meaning of this sentence is clear.

We have now changed the sentence to make it clearer.

13. P.15, last sentence: again, to correct the statement about limited supporting evidence, it might be useful to note that Hanney et al (2013) also provided detailed examples of research making an impact on health in several of the case studies.

We now report mention specifically the three case studies that talk about health gains. Please see page 16 of the revised manuscript.

14. P.16: middle para: I do not think the Hall et al study meets the inclusion criteria.

As explained earlier, we believe the study does meet the criteria.

15. P.17: several further studies also addressed the question of health gains. These include both Glover et al (2014) and Guthrie et al (2016) for whom the analysis of the rate of return is based on first establishing the level of health gain from research. Each of the 16
case studies in Hanney et al (2007) considered the question of health gain, even in those instances where it was to report that there was no evidence.

Thank you for pointing this out. Indeed, we have now included these studies in the analysis. Please see pages 16-17 of the revised manuscript.

16. P.19/20: Conclusions:

A) The Conclusion raises various important points, but in general needs to go much further in providing a nuanced analysis of the issues around the assessment of the impact of health research.

Thank you for the specific comments. The conclusion section has changed substantial and we hope that it now passes a clearer message of the main points the study is trying to make.

B) The review is correct to open up the issues around sources of bias. But, as noted, some of the account, especially around reporting bias, would benefit from considerable further analysis. Such analysis might give recognition to the extensive triangulation and presentation of evidence (especially around impact on policy) that is contained in some of the papers that have been described as failing 'to provide supporting evidence when making claims.' It is through the detailed case studies that issues such as attribution are addressed, and detailed evidence provided about the level of impact that can justifiably be claimed from specific pieces of research.

This is a fair point and we acknowledge that the paper may be giving the impression that it is too critical to the literature and does not recognise the enormous effort that this kind of work requires. In particular a number of studies used more than one methods and different data to provide evidence. We have added a paragraph in the discussion section to clarify this and recognise the cost of conducting such detailed analysis. Along with other changes that we have now made, we hope the overall message of the paper acknowledges the effort put in working in this area and encourages for more research on it.

C) One issue that could usefully be discussed is the balance between the desirability of providing the detailed evidence in case studies, such as those in Hanney et al (2007) and Guthrie et al (2015), and the resources required to do this. Aspects of this are already discussed in several of the included studies and the wider literature referenced in the text. This crucial point in the existing literature feeds into thoughts about new approaches.

Again, we absolutely agree with the reviewer. We have now put more emphasis, throughout the text (introduction, methods and now conclusions) to highlight the laborious and resource intense task of conducting assessment studies of this kind.
D) The authors are right to explore new approaches, but need to analyse them in the light of the findings of the review. So, Researchfish does provide a way of recording data on a large scale, but it is entirely self-reported and does not include the triangulation involved in the detailed case studies. It is thus open to much larger concerns about aspects of bias than arise with the case studies included in some of the papers included in the review. Similarly, the REF provides an important additional way to assess the impact of health research, but it is highly selective.

Indeed, we have added a comment highlighting the challenges of Researchfish. With respect to REF, we agree that REF is highly selective. Yet, the point we are trying to make here is that REF will push researchers to come up with more ideas of how to measure impact beyond academia, as this is increasingly becoming an even more important component of REF and therefore funding for Higher Education institutions.

E) The issue of selectivity could usefully be considered in relation to the points made at the start of the final paragraph about the degree of risk and unpredictability of research, and that impact assessment can be interpreted as neglecting the inherent value of science. Some of the assessments of the impact of research that have adopted a selective approach are partially attempting to reconcile the nature of science with the demands for evaluation. They are sometimes doing this by focusing on some of the projects that have achieved impact, but not subjecting those that have not made an impact to detailed (and potentially unnecessary?) analysis. Thus, depending on the purposes of the exercise, a more nuanced and selective approach might be both more feasible and justified in the particular circumstances of research impact assessment. Similarly, the studies by Glover et al (2014) and Guthrie et al (2016) recognise that not all research will necessarily make an impact, but take into account the costs of the entire body of research they are valuing and set the benefits from some of the research against that. In this way they are being conservative and reducing the risk that would have arisen had they compared the benefits from selective studies against the costs of just those specific studies, rather than against the costs of the whole research portfolio.

In our conclusions, we have now added an extensive discussion on how purposive sampling is the best way to show impact and celebrate successes.

F) Limitations: as noted, the authors are right to note the challenges involved in extracting precise data from all the studies. However, they should go on to recognise that in some cases these challenges arise from the very depth of evidence provided in some of the papers. This rather contradicts the authors' repeated claims about insufficient evidence being provided in some of the papers, and yet it was such claims which, in turn, had been used as the basis for the authors' categorisation of the papers as having a high risk of reporting bias. A further limitation the authors should note is that by only considering the period from 2006-17 they missed out quite a few UK peer-reviewed impact assessment studies that had been included in the 2007 review by Hanney et al. The authors
appear to try to justify this period by claiming it coincides with the creation of the NIHR, but I do not see the relevance of that as this is not a study that focuses specifically on the research funded by the NIHR.

We hope that by now we have provided more clarity about what reporting bias is and how it does not necessarily relate to the length of a report or a study.

We also acknowledge that reviewing papers from a longer period would have been great, but we effectively had to draw a line and 2006 is quite an important year, as the launch of the NIHR had a huge impact on public funding of research in the UK.

18. p.24: References: these are generally accurate, but the journal title seems to be missing from reference 16, and the publisher and location seem to be missing from reference 18.

Thank you for pointing this out. We use Zotero for our references, but wherever there was missing information we tried to correct it manually. We have now changed both, what used to be reference 16 and 18.

Reviewer #2: This is a strong, interesting and timely paper that I recommend for publication subject to the comments below. I have two significant issues that I would like to raise with the authors and a series of minor corrections/typos.

I should note at the outset that I am an author of a number of the papers that are reviewed and assessed in the study and hence have a potential conflict of interest which I did share with the editor before undertaking this review.

We are grateful to the reviewer for their positive and constructive comments. Below we explain how we address them, point by point. We do hope the reviewer finds merit in our revisions.

Significant issues

1. I am concerned about the definition used for 'selection bias' as described in Table 1. The idea that this sort of evaluative research should be based on randomisation is misguided and thus is a bias in its own right in this actual study. A lot of the studies that are reviewed use case studies where accepted cases study methodology is to use 'purposive sampling' whereby you deliberately select the project that are likely to have an impact (see Yin, R (2008) Case Study Research: Design and Methods). The issue here is not whether the projects are representative or randomised but where the sampling approach is transparently described in the research publication (which is covered in reporting bias). I think this is potentially a significant flaw in this study which biases against case study research. At a minimum this should be acknowledged and developed as a significant line of argument in the discussion, but ideally the definition of selection bias should be reviewed and re-analysed. Related but less significant is that it would be help to know the basis for the response rates - is this driven data or is a heuristic developed by the authors.
The reviewer is right in that the use of ‘randomisation’ in the definition of ‘selection bias’ can be rather misleading, and what is more, it does not apply to all types of studies we reviewed. We have therefore decided to remove it. We now clarify that selection bias refers simply to how likely the sample is to reflect the wider population from which it is selected. It aims to capture the fact that a number of assessment studies chose to analyse projects that were more likely to have an impact, therefore they were not representative of the wider pool of projects. The reviewer is also right to highlight that most of the studies that chose such purposive sampling approach are very transparent about it and mention it as a limitation of their work. We also agree that this approach may well be the best way to identify the benefits and impact of research and celebrate the successes of funders. We have modified our conclusion section on page 22 of the revised manuscript to discuss all these points.

Specifically regarding case studies, Yin (2008) acknowledges that case study methods, in particular documentation analysis, are more likely to be subject to selection and reporting bias. This does not mean though that all case study analyses are biased. In fact, our review includes studies that followed a case study approach, and scored low in the risk of selection bias, because they chose a representative sample of cases (see Hanney et al 2007), a method that Yin suggests can reduce bias.

2. The authors raise a really important issue with respect to ‘funding bias’ and one that I would like to see them expand on in the discussion. As a practitioner of this type of research it is very difficult (near impossible) to get independent funding. I and others have been arguing for this for some time, for the reasons that the authors note ie there is a potential tendency for studies to support the interests of the study, whether real or perceived. (See for example recent commentary in BMJ Open: https://bmjopen.bmj.com/content/8/9/e022131.info). It would be really helpful if the authors commented on this in the discussion and bought out the need for independent funding streams for this type of research.

We could not agree more with this comment and we would like to thank the reviewer for bringing the BMJ Open commentary to our attention. We have added a section in our conclusion discussing the importance of more independent funding streams to research on research. Please see page 21 of the revised document.

Minor comments/typos

Page 3, line 34-36: There is something wrong with the sense of this sentence (eg possible missing word)

We have now changed this sentence.

Thank you for your suggestion. We have replaced the reference with the 1996 one.

Page 4, line 32. Why these journals? I would have come up with a different list (e.g., HSRP, Research Evaluation, BMC Medicine etc.) do some explanation would be helpful Page 4, line 45.

We agree that a wider range of journals could have been used. We now acknowledge this in our limitations.

Not sure why the start of NIHR in 2006 is a good reason, but very minor comment

We felt that the establishment of the NIHR in 2006, would be a good justification for the start of our review, as it signalled the beginning of a new era for public funding for health research in the UK. Having said that, we added a comment in our limitations that we may have missed key papers published prior to 2006.

Page 20, line 24. Do you mean the use of the ResearchFish reference here? Seems wrong

It was meant to be a reference to a Welcome Trust source, which should now be corrected.