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)	Is an ounce of prevention worth a pound of cure? A cross-sectional study of the impact of English public health grant on mortality and morbidity.
AUTHORS	Martin, Stephen; Lomas, James; Claxton, Karl

VERSION 1 – REVIEW

REVIEWER	Peggy Honore Louisiana State University Health Sciences Center-New Orleans USA
REVIEW RETURNED	06-Jan-2020

GENERAL COMMENTS	This course of analysis is very important and under performed. Glad to see this move forward.

REVIEWER	Laura Vallejo-Torres Universidad de Las Palmas de Gran Canaria, Spain
REVIEW RETURNED	24-Feb-2020

GENERAL COMMENTS	The paper entitled “Is an ounce of prevention worth a pound of cure? Estimates of the impact of English public health grant on mortality and morbidity” compares the average effect of health care (treatment) expenditure and the average effect of public health (prevention) expenditure on mortality using data across 150 local areas in England. This effect is then translated into a cost per QALY estimate. The paper makes use of a similar methodology developed by the authors to estimate the average opportunity cost of the English NHS. The authors find that public health expenditure is more productive than treatment expenditure, in the range of three to four times so. The paper thus concludes that “the recent proposal to shift resources away from [NHS healthcare expenditure] and towards [public health expenditure] is an evidence-based one”. This is a well-written paper, focusing on an interesting topic that uses a robust methodology. Their findings can also have relevant implications for policy making. My main concerns are as follows: 1. It is not clear how the authors translate the estimated effect of expenditure on mortality into a cost per QALY. They appear to have used a previous estimate derived from an assumption: that the effect on morbidity is proportional to the estimated mortality effect. However, instead of applying the estimated mortality effects derived from their own calculations they used that from a previous analysis that is only based on treatment expenditure data, and used that to both public health and healthcare. This is particularly unconvincing when one sees the different effects that public health

	and treatment services have on mortality. How can the authors assume the same effect on morbidity? A more detailed explanation and justification are needed in order to understand and validate this approach. Otherwise, only mortality effects estimates should be presented. 2. It is also quite unconvincing the assumption that the effect of public health on mortality plays out in two years from the time the money is spent. And this is particularly relevant when comparing treatment and prevention care, where the lag effects are likely to be very different. The authors mention a Californian study that finds out that more than half of lives saved by public health spending occurred after two years, although they seem to have phrased that in a way that supports they approach? This is a crucial point that merits more than a few lines in the limitations section. 3. The uncertainty around the estimates appears to be very large. In fact, the 95% confidence intervals presented in the abstract overlap, showing that we could not reject the null hypothesis that the effects of public health and healthcare on mortality are the same. This is not in line with the overall confidence granted to the results in the discussion, conclusion and abstract sections of the paper. 4. The implications of using an IV approach might need to be considered. Its use implies that the estimated coefficients now have a LATE interpretation. Is it important for policy recommendations the fact that the effects on mortality estimated on this paper are only relevant for the local areas that increase/decrease expenditure due to MFF and DFT issues? 5. Finally, in more general terms, the applications of these estimates into decision making need further discussion. Do they provide evidence for allocative or productive efficiency purposes? The authors appear to suggest both. They argue that their findings provide evidence for setting a larger size of the budget to public health than currently achieved, although one can ask, how far more and how should this be established? They also mentioned that under current budget constraints new public health interventions should have a cost per QALY lower than that estimated according to their findings. Does this imply that we ought to use a different (considerably lower) cost-effectiveness threshold value for preventive rather than curative care? Is this feasible?

REVIEWER	Martin Henriksson Linköping University, Sweden
REVIEW RETURNED	30-Mar-2020

GENERAL COMMENTS	Summary This paper estimates the responsiveness of mortality to changes in healthcare and public health expenditure in the UK. The authors employ similar estimation approaches as their previous work in the field and come up with estimates of £3800 and £13500 per QALY gained of public health and healthcare spending, respectively. Major/general comments 1. The major contribution of this work is that we, for once, now can compare different broad approaches to generate population health; “prevention” and “cure”. If we have faith in these results, they carry major policy implications and I believe the authors have done some great work here.

	2. I am not familiar with the details of the UK system, and to be fair the authors are rather humble about this in the text, but it remains a bit unclear to me how much of preventative measures are included in the healthcare budget and how much "treatment" is part of the public health budget. If we could get some figures here it would give the non-initiated reader an idea of these magnitudes. 3. Another general comment is around the balance of the paper in terms of technical sophistication and the policy issue at hand. It is difficult to present such a comprehensive study with plenty of analytical choices and corners to cover and I believe the authors are doing this well although it takes some effort to cover all details and supplementary material even for a very interested reader (including myself – and I have not been through every detail of the supplementary material). Having said this, with the information provided the results and analytic methods are up for scrutiny and this is to be applauded. I would still consider the balance of the presentation and perhaps move some more of the regression methodology to the appendix and focus slightly more on the highly important policy issues raised. I also wonder whether some kind of graphical representation of the conceptual model including casual links would be a way to make the paper slightly more accessible for the general audience. Minor/technical comments 1. Abstract – methods. Are the methods for estimating QALY effects really well-established? I would argue they are probably still associated with a fair amount of uncertainty. 2. Abstract – conclusion. This comment is also valid for the conclusions in the main text to some extent. I believe the comparison with the NICE threshold may be considered a bit off target. At least it should perhaps not take up more than half of the conclusion in the abstract. I believe there is a very important policy story to be told here, preventative measures should get more funding. I would not let that conclusion half disappear in a discussion about the NICE threshold. 3. P12.L43-50: The authors should clarify that they are not doing forbidden regression, i.e., all five instruments are used in both first stages. Also, could the authors explain the three additional instruments? 4. P8.L55-P9.L5: What are the authors thoughts on health care input prices as an exogenous variable? Is price not determined by equilibrium in the market for these inputs? 5. The effect of DFT on spending seems fairly robust to specification (Table A2). What are the results (first and second stage) using only this instrument and no controls? If the results are affected by the inclusion of covariates (column 6, Table A1 and A2), why are these covariates necessary for conditional independence? 6. I am not familiar with all of the diagnostics test used by the authors. (1) Why do we care about nonlinearities (RESET test) in IV regression? (2) My understanding is that the threshold for weak instruments goes up quite drastically with more than one endogenous regressor. Do the tests used reflect this issue?

REVIEWER	Dr Gemma Bilkey Department of Health Western Australia Australia
REVIEW RETURNED	25-May-2020

GENERAL COMMENTS	Nicely written and timely work. While the final paragraph of the discussion describes discounting in a broad sense, it would be valuable to provide a comment how this may have changed the results, or a further justification for why this was not included. As an international reader, I was not clear on the remit of the CCG (is this purely tertiary spending?)

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Peggy Honore
Louisiana State University Health Sciences Center-New Orleans USA

Please leave your comments for the authors below
This course of analysis is very important and under-performed. Glad to see this move forward.

Authors' response: Thank you for your kind comments.

Reviewer: 2
Laura Vallejo-Torres
Universidad de Las Palmas de Gran Canaria, Spain

Please leave your comments for the authors below
The paper entitled "Is an ounce of prevention worth a pound of cure? Estimates of the impact of English public health grant on mortality and morbidity" compares the average effect of health care (treatment) expenditure and the average effect of public health (prevention) expenditure on mortality using data across 150 local areas in England. This effect is then translated into a cost per QALY estimate. The paper makes use of a similar methodology developed by the authors to estimate the average opportunity cost of the English NHS. The authors find that public health expenditure is more productive than treatment expenditure, in the range of three to four times so. The paper thus concludes that "the recent proposal to shift resources away from [NHS healthcare expenditure] and towards [public health expenditure] is an evidence-based one".

This is a well-written paper, focusing on an interesting topic that uses a robust methodology. Their findings can also have relevant implications for policy making.

Authors' response: Thank you for your kind comments.

My main concerns are as follows:
1. It is not clear how the authors translate the estimated effect of expenditure on mortality into a cost per QALY. They appear to have used a previous estimate derived from an assumption: that the effect on morbidity is proportional to the estimated mortality effect. However, instead of applying the estimated mortality effects derived from their own calculations they used that from a previous analysis that is only based on treatment expenditure data, and used that to both public health and healthcare. This is particularly unconvincing when one sees the different effects that public health and treatment services have on mortality. How can the authors assume the same effect on morbidity? A more

detailed explanation and justification are needed in order to understand and validate this approach. Otherwise, only mortality effects estimates should be presented.

Authors' response: We would like to thank the referee for drawing this issue to our attention and giving us the opportunity to clarify this in the paper. As suggested by the reviewer, we have added two columns to table 3 that report the cost per death averted for public health and treatment expenditure. These mortality-based estimates confirm our broader QALY-based results that public health expenditure is more productive than healthcare expenditure. The purpose of the paper is to try and demonstrate the relative health benefits of these two types of expenditure and, if possible, to compare the size of these benefits with those associated with particular types of health care expenditure (for example, on new medical technologies). To do this it is necessary to convert mortality effects into broader QALY effects. At the moment there is no evidence about the mortality effects by disease area of public health expenditure. In the absence of any evidence, we assume that the distribution of mortality benefits across disease areas for public health is similar to that for healthcare expenditure. This assumption is now made explicit on p.19 (of the revised submission) so that readers can judge for themselves the usefulness of our cost per QALY estimates. It is not obvious that this assumption will either over- or under-estimate the total QALY benefits of public health expenditure. By making this assumption we are able to compare the health effects of both types of expenditure and we can compare these effects with, for example, the NICE threshold for the adoption of new medical technologies in the NHS. Moreover, by making this (now) explicit assumption we hope to stimulate research that will examine its accuracy.

2. It is also quite unconvincing the assumption that the effect of public health on mortality plays out in two years from the time the money is spent. And this is particularly relevant when comparing treatment and prevention care, where the lag effects are likely to be very different. The authors mention a Californian study that finds out that more than half of lives saved by public health spending occurred after two years, although they seem to have phrased that in a way that supports they approach? This is a crucial point that merits more than a few lines in the limitations section.

Authors' response: In an ideal world we would have access to expenditure and mortality data that stretch back many years so that we could address this issue properly. Unfortunately such data do not exist and, instead, we make the best use of what data are available and we draw the reader's attention to the limitations that such data imply for our results. However, and as we point out in the paper, the way in which we use the expenditure and mortality data might not be as troublesome as first appears. In support of our approach we cite the Californian study that suggests over one-half of all cumulative lives saved through public health expenditure occur in the two years following that expenditure, and our mortality measure includes deaths in the expenditure year and the following two years.

Moreover, although we omit mortality effects for later years, some current mortality may reflect public health expenditure from many years ago. Implicitly we are assuming that the data represent a quasi long-run equilibrium situation, that relative expenditure levels and health outcomes within each local authority have been reasonably stable over a period of time, and that any lagged of effect of current expenditure on future mortality is offset by the impact of previous expenditure on current mortality. These are not unreasonable assumptions in the English context but they are just assumptions, and they might be less appropriate for other geographies where, for example, relative expenditure and outcomes have changed through time. We have added a few sentences to this effect on pp.23-24.

3. The uncertainty around the estimates appears to be very large. In fact, the 95% confidence intervals presented in the abstract overlap, showing that we could not reject the null hypothesis that the effects of public health and healthcare on mortality are the same. This is not in line with the overall confidence granted to the results in the discussion, conclusion and abstract sections of the paper.

Authors' response: We must thank the reviewer for highlighting this issue and enabling us to address it in the paper. Using the point and standard error estimates associated with the mortality elasticities in table 3, we undertook a simulation study of the difference between the public health and CCG QALY gains associated with the budget boost described in columns 7 and 8 of table 3. We made one million pairs of draws from the two distributions. We found that the size of the public health QALY gain was greater than the size of the CCG QALY gain in just over 94% of the draws from the backward selection estimates, and that this proportion increased to over 99% when the forward selection estimate were used. We feel that this allows us to conclude that the public health QALY effect is greater than the CCG effect. We have added details of this simulation to the paper on p.20.

4. The implications of using an IV approach might need to be considered. Its use implies that the estimated coefficients now have a LATE interpretation. Is it important for policy recommendations the fact that the effects on mortality estimated on this paper are only relevant for the local areas that increase/decrease expenditure due to MFF and DFT issues?

Authors' response: We believe this is unlikely to be great policy significance. The DFT and MFF variables are not binaries but are continuous variables that affect all LAs in their 2013/14 allocations. Moreover, we obtained similar cost per QALY results for CCG expenditure using an entirely different set of instruments a few years ago (see https://www.ncbi.nlm.nih.gov/books/NBK274315/ for details). Consequently, we feel that our estimated LATE is generalizable to other situations where CCGs/LAs experience budgetary changes (e.g., when discretionary expenditure is reduced through the commitment to a new specific technology).

5a. Finally, in more general terms, the applications of these estimates into decision making need further discussion. Do they provide evidence for allocative or productive efficiency purposes? The authors appear to suggest both. They argue that their findings provide evidence for setting a larger size of the budget to public health than currently achieved, although one can ask, how far more and how should this be established?

Authors' response: Our findings provide evidence for allocative efficiency purposes. Our study is motivated by the UK government's proposal to stop cutting the local authority public health grant by re-allocating part of the treatment budget to preventative activity. Our results suggest that this proposal is an evidence-based one although our estimates and other evidence also suggest that an increase in CCG expenditure would provide good value for money too (eg when compared with the Treasury's estimate of the consumption value of health). Our results do not allow us to recommend the size of the increase in the public health budget. However, given the very low PH marginal cost per QALY relative to that for CCG expenditure and to the NICE threshold for the adoption of new technologies by the NHS, we would recommend a return to pre-austerity expenditure levels.

5b.They also mentioned that under current budget constraints new public health interventions should have a cost per QALY lower than that estimated according to their findings. Does this imply that we ought to use a different (considerably lower) cost-effectiveness threshold value for preventive rather than curative care? Is this feasible?

Authors' response: If the objective is to maximise total health gain then, conditional on the current allocation of PH and NHS expenditure, our results suggest that the CET for interventions funded from PH resources should be lower than those funded from the NHS budget. Of course this also suggests that reallocating resources from NHS to PH would improve health until the CET for PH and NHS was equalised. Our results also suggest that the NICE threshold for all new technologies, whether drawing on the NHS budget or the PH grant, should be lower than the current CET.

Reviewer: 3

Reviewer Name
Martin Henriksson

Institution and Country
Linköping University, Sweden

Please leave your comments for the authors below
Summary
This paper estimates the responsiveness of mortality to changes in healthcare and public health expenditure in the UK. The authors employ similar estimation approaches as their previous work in the field and come up with estimates of £3800 and £13500 per QALY gained of public health and healthcare spending, respectively.

Major/general comments
1. The major contribution of this work is that we, for once, now can compare different broad approaches to generate population health; "prevention" and "cure". If we have faith in these results, they carry major policy implications and I believe the authors have done some great work here.

Authors' response: Thank you for your kind comments.

2. I am not familiar with the details of the UK system, and to be fair the authors are rather humble about this in the text, but it remains a bit unclear to me how much of preventative measures are included in the healthcare budget and how much "treatment" is part of the public health budget. If we could get some figures here it would give the non-initiated reader an idea of these magnitudes.

Authors' response: Precise figures of the break down between prevention and treatment within the PH grant and NHS budget are not available. As one very rough guide to the volume of preventative expenditure within the treatment total, CCG programme budgeting data for 2013/14 reports a total spend of £65bn of which £411m (less than 1%) is in the 'Healthy Individuals' programme and could be described as for preventative activity. With regard to the public health grant, there is the issue about how to view treatment expenditure that also has a preventative effect. For example, of the £2.5bn public health grant about £489m was spent on drug and alcohol misuse, and £381m on STI testing/treatment. So, if we ignore the preventative element associated with these expenditure components, it could be argued that up to £870m (35%) of the public health grant is on treatment. But, of course, part of this expenditure will have a preventative effect too. This issue is acknowledged in section 2.1 with further discussion in appendix section A1. Strictly speaking, we are comparing the productivity of the public health grant with CCG healthcare expenditure. However, we believe that it is reasonable to think of this as a comparison of the marginal productivity of preventative and treatment expenditure although our primary purpose is to estimate the marginal effect of these two different sources of public expenditure which are subject to different budgetary constraints/choices.

3. Another general comment is around the balance of the paper in terms of technical sophistication and the policy issue at hand. It is difficult to present such a comprehensive study with plenty of analytical choices and corners to cover and I believe the authors are doing this well although it takes some effort to cover all details and supplementary material even for a very interested reader (including myself – and I have not been through every detail of the supplementary material). Having said this, with the information provided the results and analytic methods are up for scrutiny and this is to be applauded. I would still consider the balance of the presentation and perhaps move some more of the regression methodology to the appendix and focus slightly more on the highly important policy issues raised. I also wonder whether some kind of graphical representation of the conceptual model

including casual links would be a way to make the paper slightly more accessible for the general audience.

Authors' response: Thank you for your kind comments. We agree that it is very difficult to present this material so that it is both robust enough for the academic audience yet reasonably easy to follow for the more policy orientated reader. A good deal of the regression methodology is already in the appendix and we are reluctant to move more. However, we think that your suggestion to add a graphical representation of the conceptual model including casual links as a way of making the paper slightly more accessible for the general audience is an excellent one. Hence we have added a new figure and an additional explanatory paragraph of text to address this issue on p.9.

Minor/technical comments

1. Abstract – methods. Are the methods for estimating QALY effects really well-established? I would argue they are probably still associated with a fair amount of uncertainty.

Authors' response: These methods have been around for a few years now and have been used in several studies. We are not aware of any major criticisms or better alternatives given the data available so we are reasonably happy with them. Moreover, the very recent paper by Soares, Sculpher and Claxton (2020) presents an application of the structured elicitation of the judgments of key individuals (including clinical experts) about the size of the QALY benefits associated with English healthcare expenditure. This study, available at https://journals.sagepub.com/doi/abs/10.1177/0272989X20916450?journalCode=mdma, finds that although most experts found replying to the questions challenging, they were able to express their beliefs quantitatively. The experts' judgements suggest that the assumptions made by earlier work that estimated the quality-adjusted life-year (QALY) impacts of changes in expenditure are likely to have underestimated the QALY benefits and, as a consequence, to have overestimated the "central" estimate of the health opportunity cost associated with NHS expenditure (£12,936 per QALY).

2. Abstract – conclusion. This comment is also valid for the conclusions in the main text to some extent. I believe the comparison with the NICE threshold may be considered a bit off target. At least it should perhaps not take up more than half of the conclusion in the abstract. I believe there is a very important policy story to be told here, preventative measures should get more funding. I would not let that conclusion half disappear in a discussion about the NICE threshold.

Authors' response: We believe this paper provides evidence which can inform resource allocation and decisions across these two categories of public expenditure, which includes decisions made by NICE which carry a funding mandate (approved interventions must be funded). We are not convinced that a comparison of our results with the NICE threshold dilutes the finding that PH expenditure is more productive of health than NHS expenditure (at the margin). We feel that it is important to draw attention to just how much more productive of health both types of expenditure are than the threshold currently used by NICE.

3. P12.L43-50: The authors should clarify that they are not doing forbidden regression, i.e., all five instruments are used in both first stages. Also, could the authors explain the three additional instruments?

Authors' response: Our understanding of forbidden regression in the IV context comes from section 4.6.1 of Angrist and Pischke's book 'Mostly Harmless Econometrics: An Empiricist's Companion'. This focuses on such issues as how to handle dummy instruments and non-linearities in the first-stage, and the importance of including the same group of covariates in both the first and second stages. We are not convinced that the inclusion of all five instruments in both first stages is 'forbidden' regression.

We use Stata to estimate our specifications and we are unaware of how to estimate the specifications without including all 5 instruments in both first stages. We start by estimating the 'full' specification (i.e., with all controls and all instruments included) whether we are estimating a public health only or public health and treatment expenditure regression. We then use backward or forward selection to eliminate irrelevant controls and/or problematic instruments. The three additional instruments are for CCG expenditure and are explained in appendix A3. They are very similar to those for public health expenditure but relate to the allocation of CCG budgets rather than the public health budget. They comprise: the distance from the target allocation; the market forces factor; and the prescribing cost age index.

4. P8.L55-P9.L5: What are the authors thoughts on health care input prices as an exogenous variable? Is price not determined by equilibrium in the market for these inputs?

Authors' response: This instrument is suggested by the funding rule approach and, as always, we are guided by the Hansen-Sargan test for instrument validity. The local input price index could be correlated with unmeasured determinants of mortality but conditionally exogenous variation is likely to remain once other controls for need are used since the price index is unlikely to be a perfect adjustment. We have over a dozen potential socio-economic covariates (including the Index of Multiple Deprivation) in the full specification mortality equation and hence it is difficult to imagine what deprivation effect the input price index would detect that our covariates do not. Moreover, both MFFs (for public health and treatment expenditure) are not included as instruments in our preferred backward and forward parsimonious specifications so the issue does not really arise with our final results.

5. The effect of DFT on spending seems fairly robust to specification (Table A2). What are the results (first and second stage) using only this instrument and no controls? If the results are affected by the inclusion of covariates (column 6, Table A1 and A2), why are these covariates necessary for conditional independence?

Authors' response: We agree that the effect of DFT on spending seems fairly robust to the precise specification (Table A2). If we re-estimate the specification in column 6 of tables A1 and A2 (i.e., mortality as a function of spend with no controls, and DFT is the only instrument) then the coefficient on expenditure in the second-stage is +0.133 [t-ratio=1.97] and in the first-stage equation the coefficient on DFT is +1.178 [t-ratio=9.77]. Without the controls for need we detect a positive association between spend and mortality rather than the causal effect of expenditure on outcome. We require the controls for health care need to make the instruments conditionally exogenous.

6. I am not familiar with all of the diagnostics test used by the authors. (1) Why do we care about nonlinearities (RESET test) in IV regression? (2) My understanding is that the threshold for weak instruments goes up quite drastically with more than one endogenous regressor. Do the tests used reflect this issue?

Authors' response: (1) We care about nonlinearities because they suggest an omitted effect which, if ignored, might result in inconsistent coefficient estimates. (2) We use the Sanderson-Windmeijer test for the strength of the instruments associated with each individual endogenous regressor. These statistical results are generated as part of the output associated with the ivreg2 routine in Stata and are specifically designed for the presence of more than one endogenous regressor. We understand that there is no widely accepted rule of thumb threshold for weak instruments when there are two endogenous instruments. In the absence of theoretical guidance, we persevere with the single instrument rule of thumb and report the relevant test statistic so that the reader can judge for themselves. Moreover, the Sanderson-Windmeijer F-statistic for the public health instrument is way

above ten in both the preferred backward and forward selection specifications (it is 70.8 in the former and 57.0 in the latter).

Reviewer: 4
Dr Gemma Bilkey
Department of Health
Western Australia
Australia

Please leave your comments for the authors below
Nicely written and timely work. While the final paragraph of the discussion describes discounting in a broad sense, it would be valuable to provide a comment how this may have changed the results, or a further justification for why this was not included.

Authors' response:
The referee is quite right that we do not discuss discounting of the QALY effects of PH and NHS expenditure. We now report cost per death averted, which does not require discounting and cost per QALY in Table 3 (see our responses to referee 2).

The translation of the estimated mortality effects to QALY effects is based on previous work which also used estimated mortality effects of changes in NHS expenditure to calculate the QALY effects. In this previous work the reported estimates reflect changes in undiscounted QALYs associated with changes in expenditure. Discounting these quality adjusted life year effects in previous work at 3.5% led to a very modest increase the cost per QALY (from £12,936 to £13,141 in Claxton et al 2015 (see https://www.ncbi.nlm.nih.gov/books/NBK274315/). The effects of discounting are modest because the health effects of changes in expenditure are restricted to one year. A large proportion of this health effect is quality of life (which occurs in that year so is not subject to discounting). The change in mortality due to a change in spend that occurs in that year does have life year effects (adjusted for quality) in subsequent years which are subject to discounting. Some changes in mortality will have life year effects over many years and other mortality effects will not. On average 4.5 life years is associated with each death averted, so, on average, the effect of discounting is modest even when a rate of 3.5% rate is applied, when 1.5% or lower is arguably more appropriate for health.

Since discounting future health effects would apply equally to the effects of PH and NHS expenditure it would not change the comparison of the effects of PH or NHS expenditure. Of course, should more waves of data make it possible to estimate a longer lag structure then discounting would become more important (see our responses to reviewer 2 on these issues). However, overall this is likely to capture more total discounted health effects of changes in expenditure, reducing rather than increasing the estimates of cost per QALY for both PH and NHS expenditure. If the effects of PH expenditure tend to have longer lags than NHS expenditure then the total albeit discounted effects would tend to be greater for PH, reinforcing the findings of this paper.

For all these reasons we would on balance prefer to avoid the complications of a full discussion of discounting in the current text as doing this issue justice would not be feasible in this paper and failing to fully explain the issues is likely to confuse most readers who may not be familiar with debate about why health effects should be discounted and what an appropriate discount rate for health should be.

As an international reader, I was not clear on the remit of the CCG (is this purely tertiary spending?)

Authors' response: To illuminate this issue we have added some words to the end of the final sentence on p6 so that it now reads as:

For our study year (2013/14), each authority (CCG) was assigned a fixed annual budget by the national ministry (the Department of Health) within which they were supposed to meet expenditure on most types of health care including inpatient care, outpatient and community care, and pharmaceutical prescriptions.

VERSION 2 – REVIEW

REVIEWER	Laura Vallejo Torres Universidad de Las Palmas de Gran Canaria, Spain
REVIEW RETURNED	28-Jul-2020

GENERAL COMMENTS	The clarifications and extra analyses undertaken by the authors have addressed by previous concerns.