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)	Effect of a national requirement to introduce named accountable general practitioners for patients aged 75 or older in England: regression discontinuity analysis of general practice utilisation and continuity of care
AUTHORS	Barker, Isaac; Lloyd, Therese; Steventon, Adam

VERSION 1 - REVIEW

REVIEWER	Mark Wilberforce University of Manchester, England
REVIEW RETURNED	02-Mar-2016

GENERAL COMMENTS	The manuscript contains a well-executed quasi-experimental study of whether a named GP leads to improved care continuity. A regression discontinuity design is appropriate to this policy change. Such designs are ideal for situations where policy eligibility is based on arbitrary criteria (such as age cut-offs, here) allowing some of the benefits of an RCT to be mimicked. The paper is drafted in an accessible manner and contains the majority of the information needed for readers to critically appraise its content. I have a few thoughts which I feel would improve it further. My main concern relates to the intervention being evaluated itself and whether it merits the research attention. The introduction and/or methods need to establish a firmer rationale for this being an interesting and valuable addition to knowledge which would inform future policy/practice. Specifically, I am drawn to an incongruity. On the one hand, the wider policy is described as being a vehicle for GPs taking responsibility for coordinating all health and social care needs, and empowering patients to demand greater continuity of care, with importance illustrated by (for example) reporting meta-analysis of case management. Yet on the other hand, the implementation is simply in the form of a standardised unsolicited letter, telling patients the name of their GP. Perhaps the cynic in me sees this policy as political gimmickry, but I wonder if GPs or patients would genuinely expected any changes to come about as a result of this letter. I am not aware of much interest in this policy among primary care professionals. I also note for the wider "roll-out" of this initiative to other age groups, they have dropped the insistence that a letter is sent to patients, so it is entirely an administrative change. In revisions, I would urge the authors to reflect on the rationale for exploring the policy and want they genuinely would expect to see from its implementation. My other comments are more specific, and are as follows: * I feel that the researchers should think again about their primary

	outcome. Based on everything in the introduction and methods up to this point of the manuscript (p8), the primary outcome must be the usual provider of care (UPC) index? The primary outcome shouldn't be the variable that gives the best power - it should be what is most relevant to the research aim. The power calculation should be for this outcome, even if it is lower. * Further to that observation, I would also welcome a brief indication of the distribution of the UPC index at baseline. For how many patients was the UPC index already at or close to 1.0, thus facing a ceiling effect? * I also expected the limitations section to include the fact that the the analysis is blind to whether consultations (UPC, # contacts etc) were actually with the patients' named GP. It is clearly stated earlier, but should be considered in discussion. * Finally, there are a few drafting issues to attend to. E.g. opening paragraph makes a bold statement about overhaul of primary care structures definitively causing a 4 point change in continuity of care. A quick scan may pick up a few others too. I hope these thoughts are helpful to the authors.

REVIEWER	Ioannis Bakolis Department of Biostatistics King's College London Institute of Psychiatry, Psychology and Neuroscience
REVIEW RETURNED	24-Mar-2016

GENERAL COMMENTS	The authors aim to evaluate whether the introduction of named accountable GPs nationally in England was associated with improvements in longitudinal continuity of care at age 75.The authors employed a quasi-experimental regression discontinuity design. 1) Guidelines for the presentation of RD designs (Imbens & Lemieux 2008; Lee & Lemieux 2008; Bor et al 2014; Moscoe et al. 2015) make several recommendations to achieve clarity and support the validity of the design. In particular, RD validity depends on whether the data generating process is what the authors say it is. We need pictures to believe this. I strongly encourage the authors to follow these recommendations as they make it very easy for a ready to judge the quality of the design and see the effect. a. From Figure 1 in the supplementary material in seems to be a "bunching" effect in the histogram for the assignment variable which could signal manipulation. b. From Figure 2 in the supplementary material it does not seem to be balance in covariates at the threshold by plotting average values of covariates against the binned assignment variable. c. Plot all your outcomes against the binned assignment variable. Do this for each outcome. Show us the discontinuity. My main point is that if there is not a graphical validity for an RD design why the authors employed the RD method?

	2) I think it would strengthen their case (at least as a sensitivity analysis) to take a simpler RD approach and to really try to model outcomes right around the threshold. 3) Semantics: There is a lot of confusion in the RD literature on the proper use of the term "fuzzy RD". It is really an unfortunate name. "Fuzzy RD" occurs when not all subjects are affected by the threshold rule, and therefore the effect of the threshold rule should be interpreted as an "intent-to-treat" and complier causal effects can be recovered by instrumental variables techniques. Fuzzy RD is NOT when there was some cheating or phased roll-out in the area just around the threshold. 4) Authors that RD is a quasi-experimental design, similar to randomized experiment and could replicate the results. I would downplay this sentence. Statements like "similar to RCTs" are dangerous unless justified on the particular case. RD exploits a threshold rule data-generating process which creates comparable populations with different exposure statuses just above and below a threshold. (By design, that is, not by assumption). 5) Page 3 remove “causal” from evaluation 6) I would omit second sentence in page 3. Regression discontinuity is a quasi – experimental design. If there is observed confounding the models should be adjusted for. Still unmeasured trends could influence the RD models.

VERSION 1 – AUTHOR RESPONSE

Reviewer's Comment: The manuscript contains a well-executed quasi-experimental study of whether a named GP leads to improved care continuity. A regression discontinuity design is appropriate to this policy change. Such designs are ideal for situations where policy eligibility is based on arbitrary criteria (such as age cut-offs, here) allowing some of the benefits of an RCT to be mimicked. The paper is drafted in an accessible manner and contains the majority of the information needed for readers to critically appraise its content. I have a few thoughts which I feel would improve it further.

Author’s reply: We thank the reviewer for their comments.

Reviewer's Comment: My main concern relates to the intervention being evaluated itself and whether it merits the research attention. The introduction and/or methods need to establish a firmer rationale for this being an interesting and valuable addition to knowledge which would inform future policy/practice. Specifically, I am drawn to an incongruity. On the one hand, the wider policy is described as being a vehicle for GPs taking responsibility for coordinating all health and social care needs, and empowering patients to demand greater continuity of care, with importance illustrated by (for example) reporting meta-analysis of case management. Yet on the other hand, the implementation is simply in the form of a standardised unsolicited letter, telling patients the name of their GP. Perhaps the cynic in me sees this policy as political gimmickry, but I wonder if GPs or patients would genuinely expected any changes to come about as a result of this letter. I am not aware of much interest in this policy among primary care professionals. I also note for the wider "roll-out" of this initiative to other age groups, they have dropped the insistence that a letter is sent to patients, so it is entirely an administrative change. In revisions, I would urge the authors to reflect on the rationale for exploring the policy and want they genuinely would expect to see from its

implementation.

Author's reply: It is worth making a distinction between the likely effectiveness of an intervention and whether it merits research attention. There are several arguments for conducting this study; not least the £290 million of taxpayer money was spent on these approaches in 2014/5 alone. The requirement for named accountable GPs was broadened to cover all ages in 2015/6, while also remaining in the GP contract for 2016/7. This roll out has had an opportunity cost as well as imposing an administrative burden on practices – an evaluation is needed to provide objective information about whether this was worth it.
We reject strongly the notion that a study should not be published because the findings were negative. And while we value the perspective of the reviewer, we are also concerned at his implication that significant investments in health care should not be evaluated because of political intent. Surely there is a greater, rather than lesser, need for objective evaluation in those instances.
The reviewer says that he is 'not aware of much interest in this policy among primary care professionals' but every general practice in England has been required to take part in it. In addition to interest from primary care professionals, we believe the study will be of interest to a research audience because we provide a model for evaluation using routine data and regression discontinuity designs – a method which has been underutilised in clinical and health policy research, as highlighted by a recent publication in the BMJ (Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. Bmj 2016;:i1216. doi:10.1136/bmj.i1216). Our study protocol has already been published in International Journal of Integrated Care, reflecting the research interest in this topic.
We have made some minor changes to the wording of the manuscript, to better highlight why this evaluation is necessary and of interest, and to highlight some recent literature and policy changes of interest.

Reviewer's Comment: I feel that the researchers should think again about their primary outcome. Based on everything in the introduction and methods up to this point of the manuscript (p8), the primary outcome must be the usual provider of care (UPC) index? The primary outcome shouldn't be the variable that gives the best power - it should be what is most relevant to the research aim. The power calculation should be for this outcome, even if it is lower.

Author's reply: We agree with the reviewer that the primary outcome should be the one most relevant to the research aim, and we have clarified in the manuscript (page 9) that the ultimate policy objective was to improve patient outcomes while reducing cost, particularly through reduced hospital admissions. Thus, in this study we examined several intermediate outcomes (UPC index, GP contact, referrals, etc.). This is common research practice – for example, many studies of antihypertensive drugs are not able to examine cardiovascular events.
We have preferred to retain GP contacts as the primary endpoint for a mix of theoretical, pragmatic and research governance reasons. First, although we understand the reviewer's logic, as we say in the paper, we consider UPC to be upstream of the other endpoints in the logic model for the intervention (for example, referrals to specialist care might be influenced by the level of continuity achieved). This motivates the choice of other endpoints as primary. Second, the UPC index is only defined for the subgroup of study participants with at least two general practitioner contacts during the study period. Third, the sample size for the study is so large that these considerations are unlikely to make a difference to the study conclusions. Finally, in the published study protocol, we stated that the number of GP contacts would be our primary outcome - we preferred to carry out the evaluation in line with the published protocol wherever possible.

Reviewer's Comment: Further to that observation, I would also welcome a brief indication of the

distribution of the UPC index at baseline. For how many patients was the UPC index already at or close to 1.0, thus facing a ceiling effect?

Author's reply: We thank the reviewer for this suggestion and we have included the distribution in the supplementary material, and addressed it in the discussion. From visual appraisal of the distribution, we acknowledge that the reviewers comment on a ceiling effect is a slight cause for concern; however we do not believe it is a major issue given the nature of the treatment effect we are estimating.

Reviewer's Comment: I also expected the limitations section to include the fact that the analysis is blind to whether consultations (UPC, # contacts etc) were actually with the patients' named GP. It is clearly stated earlier, but should be considered in discussion.

Author's reply: This observation is appreciated, and we have added to the discussion to highlight this limitation – however, we note that it is critical in evaluations to define outcomes in the same way between treated and untreated patients. Untreated patients did not receive a named GP and so we could not define the outcome in the way the reviewer seems to be suggesting without causing bias.

Reviewer's Comment: Finally, there are a few drafting issues to attend to. E.g. opening paragraph makes a bold statement about overhaul of primary care structures definitively causing a 4 point change in continuity of care. A quick scan may pick up a few others too.

Author's reply: We have changed the wording, so that it is not suggested that the overhaul of primary care structures has definitively caused a decrease in continuity of care.

Reviewer's Comment: 1) Guidelines for the presentation of RD designs (Imbens & Lemieux 2008; Lee & Lemieux 2008; Bor et al 2014; Moscoe et al. 2015) make several recommendations to achieve clarity and support the validity of the design. In particular, RD validity depends on whether the data generating process is what the authors say it is. We need pictures to believe this. I strongly encourage the authors to follow these recommendations as they make it very easy for a ready to judge the quality of the design and see the effect.
a. From Figure 1 in the supplementary material in seems to be a "bunching" effect in the histogram for the assignment variable which could signal manipulation.
b. From Figure 2 in the supplementary material it does not seem to be balance in covariates at the threshold by plotting average values of covariates against the binned assignment variable.
c. Plot all your outcomes against the binned assignment variable. Do this for each outcome. Show us the discontinuity.
My main point is that if there is not a graphical validity for an RD design why the authors employed the RD method?

Author's reply: We thank the reviewer for these references – we consulted many of these during the design and conduct of the study and can confirm that we followed best practice in this area. Below, we address the specific points raised.
a. While the reviewer is correct that it is important to check for signs of manipulation in the running variable, this seems very unlikely to have happened in this instance – this would mean that patients or doctors have falsified age in medical records in order to influence the assignment of named GPs. We do not think this is plausible, but in any case have presented the requisite information in the supplementary material. While the number of people per age group is not uniform, , we believe the changes between the ages of 65 and 70 relate to an increase in birth rate immediately following the end of the Second World War. Thus, this check, although important to do, does not signal a problem

for the analysis.
b. We believe that the bottom four graphs (g-j) of supplementary material 2 do show balance in the covariates across the threshold, which, coupled with the large treatment jump shown in figure 2, would lead us to be confident that any discontinuity in outcomes estimated can be attributed to the jump in treatment, rather than any discontinuity in covariates.
c. We thank the reviewer for this comment, as it highlighted how Supplementary Material 2 might have been unclear. We had plotted each outcome against the binned assignment variable, as can be seen in graph a-f of Supplementary Material 2. We have changed the title of the figure to make it clearer.
Following guidelines for regression discontinuity validity checks (Imbens & Lemieux 2008), given the large treatment jump shown in figure 2, the smoothness of the assignment variable across the threshold shown in Supplementary material 1 and the smoothness of the covariates across the threshold shown in Supplementary Material 2 (graphs g-j), we believe that there is sufficient evidence that a RD design is valid.

Reviewer's Comment: 2) I think it would strengthen their case (at least as a sensitivity analysis) to take a simpler RD approach and to really try to model outcomes right around the threshold.

Author's reply: While we agree that simpler is generally better, in this particular study, the approach suggested is not possible. As highlighted in the methods section, one limitation, common to most analyses of routine data, is that we lacked information on full date of birth. As a result ,the assignment variable was discretised and we cannot tell if those born in 1939 were intended to receive treatment or not, meaning we cannot model the outcomes right around the threshold. Although we could not examine changes in outcomes very close to the threshold, we did carry out a very wide range of sensitivity analysis and this consistently showed that the intervention had no clinically significant effects on the study outcomes.

Reviewer's Comment: 3) Semantics: There is a lot of confusion in the RD literature on the proper use of the term "fuzzy RD". It is really an unfortunate name. "Fuzzy RD" occurs when not all subjects are affected by the threshold rule, and therefore the effect of the threshold rule should be interpreted as an "intent-to-treat" and complier causal effects can be recovered by instrumental variables techniques. Fuzzy RD is NOT when there was some cheating or phased roll-out in the area just around the threshold.

Author's reply: We note the reviewer's comments and agree that there is confusion in the RD literature surrounding fuzzy designs. We use 'fuzzy design' in the precise sense described by Imbens and Lemieux in their 2008 paper "Regression discontinuity designs: A guide to practice". This paper is cited in our bibliography and recommends fuzzy designs in settings such as ours when there is non-compliance. We have described this method in the paper and emphasise that the treatment effects apply to compliers aged 75. As mentioned above, we do not think that patients or doctors 'cheated' by misreporting age to influence the assignment of named GPs. And while there was phased roll out, for the period investigated eligibility for named GPs was restricted to the over 75 group. We hope this outlines our thinking and satisfies the reviewer.

Reviewer's Comment: 4) Authors that RD is a quasi-experimental design, similar to randomized experiment and could replicate the results. I would downplay this sentence. Statements like "similar to RCTs" are dangerous unless justified on the particular case. RD exploits a threshold rule data-generating process which creates comparable populations with different exposure statuses just above and below a threshold. (By design, that is, not by assumption).

Author's reply: We appreciate the reviewer's comment and agree that the sentence about being "similar to RCTs" was imprecise. We have changed the language to better reflect the sentence (see page 10).

Reviewer's Comment: 5) Page 3 remove "causal" from evaluation

Author's reply: We have removed "causal", as requested (see page 3)

Reviewer's Comment: 6) I would omit second sentence in page 3. Regression discontinuity is a quasi – experimental design. If there is observed confounding the models should be adjusted for. Still unmeasured trends could influence the RD models.

Author's reply: We have changed the second sentence on page 3, and believe this will addresses the reviewer's point, which is well taken.

VERSION 2 – REVIEW

REVIEWER	Mark Wilberforce University of Manchester, UK
REVIEW RETURNED	11-May-2016

GENERAL COMMENTS	Thank you for giving me opportunity to review the revised manuscript. The authors have given a full reply to my earlier comments, which I'm grateful for. I suggested that an improved case was needed to justify that this intervention was of real substance. Why does it warrant their (very solid) analysis? I think the authors feel that it should be self-evident to the reader. I don't want to appear obtuse, but their stiff defence didn't convince: * Although the policy has been "rolled-out", again, what does that actually mean in practice? I understand that 'roll-out' has dropped any requirement for surgeries to notify patients of their named GP. In my own surgery, the policy amounts to posting an A4 notice on the noticeboard that patients can ask at reception if they wish to know. * We don't know the economic cost of the policy: the £290m is the increase in primary care core funding but how much is spent on this policy could be very small * Yes, practices had to do it, but then practice managers are under an obligation to fulfil many duties under contractual arrangements, but does this translate to real interest? I haven't seen any editorials, letters or studies referencing this policy...? Please also note that I did not suggest that studies with negative results should not be published. I also have no objection to the political (or otherwise) origins of policies – of course – where they are of substance. In the end it's an editorial decision as to whether this is important enough, so I will leave to their judgement.

	On other issues - I'm grateful for the overview of the intermediate vs primary outcomes issue. It's a bit of a messy logic process, but I get where they are coming from. I am also glad for the distribution of UPC scores. The ceiling effect looks quite pronounced to me? A rule of thumb I was given is that a ceiling poses a problem where it exceeds 15% of the sample. That's obviously arbitraty, and I suppose this is borderline on that basis. Still, I wouldn't be quite as dismissive.

REVIEWER	Ioannis Bakolis Department of Biostatistics Department of Health Services and Population Research King's College London Institute of Psychiatry, Psychology and Neuroscience, De Crespigny Park, London SE5 8AF
REVIEW RETURNED	25-May-2016

GENERAL COMMENTS	I am satisfied with the answers of the authors and their efforts to address my comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer's comment:
* Although the policy has been "rolled-out", again, what does that actually mean in practice? I understand that 'roll-out' has dropped any requirement for surgeries to notify patients of their named GP. In my own surgery, the policy amounts to posting an A4 notice on the noticeboard that patients can ask at reception if they wish to know.
Author's reply:
We have clarified on page 14 that, following the end of our study, the requirement to write to inform patients of their named accountable GP was removed.

Reviewer's comment:
* We don't know the economic cost of the policy: the £290m is the increase in primary care core funding but how much is spent on this policy could be very small
Author's reply:
We accept that the economic cost of the intervention is unknown and have made this clearer on page 14.
We have changed the wording in the introduction (page 4), relating to the funding of the policy.

Reviewer's comment:
* Yes, practices had to do it, but then practice managers are under an obligation to fulfil many duties under contractual arrangements, but does this translate to real interest? I haven't seen any editorials, letters or studies referencing this policy...?
Author's reply:
We agree with him that this is an editorial decision. It is not our intention to refute his claim that for some general practices in England, this was a tick-box exercise. But we conducted this study because we believe it is important to evaluate changes to national policy.

Reviewer's comment:
On other issues - I'm grateful for the overview of the intermediate vs primary outcomes issue. It's a bit of a messy logic process, but I get where they are coming from.

I am also glad for the distribution of UPC scores. The ceiling effect looks quite pronounced to me? A rule of thumb I was given is that a ceiling poses a problem where it exceeds 15% of the sample. That's obviously arbitraty, and I suppose this is borderline on that basis. Still, I wouldn't be quite as dismissive.

Author's reply:

We have changed the wording of the limitation mentioning a ceiling effect on page 14, to acknowledge that our chosen metric of continuity of care (the UPC index) had an upper bound of 1 and 15% of patients reached this maximum.