PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	What are the costs associated with child and maternal health care within Australia? A study protocol for the use of data linkage to identify health service use, and health system and patient costs.
AUTHORS	Callander, Emily; Fox, Haylee

VERSION 1 – REVIEW

REVIEWER	Assoc Professor Georgina Chambers National Perinatal Epidemiology and Statistics Unit, University of New South Wales, Australia
REVIEW RETURNED	22-Jun-2017

GENERAL COMMENTS	Protocol paper such as these are useful for other researchers, but also for verify that the finding a reported relative to a-prior aims and hypotheses. I have indicated publication after revisions. My main concern with the paper is that it doesn't describe the methods fully enough, particularly for those who aren't familiar with data linkage and the Australian datasets being used. Protocol paper can afford to go into greater details of the methods proposed. These are my suggestions and comments as I systematically read through the paper: Introduction The introduction is relevant, but it may be worth giving more context for an international journal (e.g. Australian healthcare system, population of QLD, number babies, % Indigenous etc) Suggest not splitting Primary and secondary aims Data I would like to see a diagram here of the linkages and the years, plus the non-linked survey data. I think this would be very valuable, and make the paper more interesting and easy to follow. For each data set provide a link for people if they want more information I think more detail is needed on how the linkage will actually work, and in what environment it will be analysed (e.g. MBS and PBS have strict requirements). I would also like a formal description of each dataset, and who is Data Custodian More details on HHS funding and Costing Unit Data needed. For the AHS, what does 'similar demographic characterises' mean. Are these for women giving birth?

	I am confused about the time frames for the data linkage and what data will be provided. Mothers and Babies will be identified from the PDC for 2012-2015, but outcomes will be collected to 5 years – how is this possible? Also, pregnancy data is also being collected, but I'm not sure how this will occur either, more details needed. Hence the need for a diagram. Furthermore, please check how you will manage the >5 years of Medicare data needed (or are your censoring mother and babies). It an overstatement that you will obtain a 'full range of health system costs' (as you later point out). Outpatient costs are missing also. What 'clinical outcome measures' will be captured (line 36)? Administrative data isn't very good for this, so examples would be useful Analysis Please explain, why linear regression is being used to stratify by different SES and Geo regions. I did not understand this. The last 6 paragraphs of Analysis are replications with only one word change – suggestion condensing. Because this is a cost model, what will be used for discounting? Limitations What does 'once caps reached' mean? What about cross jurisdiction flows.

REVIEWER	Sanjay K Mohanty International Institute for Population Sciences (IIPS), India
REVIEW RETURNED	28-Jun-2017

GENERAL COMMENTS	I am not sure whether it is appropriate to publish a study design that will be executed in later stage. Ideally, results/design are presented in scientific journal. Review is adequate and conceptualisation is strong. May be good to present a systematic review on health disparities in Australia

REVIEWER	Tadeja Gracner Department of Economics University of California, Berkeley
REVIEW RETURNED	21-Aug-2017

GENERAL COMMENTS	Summary: This paper describes an analysis plan to quantify health service use and out-of-pocket healthcare expenditure associated with childbearing and early childhood, using several sources of administrative data from Queensland, Australia. Authors also provide ethics and dissemination plan. Comments: In the Introduction, the authors present extensive evidence on how limited access to care may impact health outcomes. However, there is little information on the current status of out-of pocket costs overall and by different socio-economic groups. I suggest that the authors spend some time describing statistics on socio-economic disparities in access to care and related health care costs disparities (even if unrelated to perinatal care) in both, Australia and Queensland. That way one can get a sense on severity of the problem, as well as how different or not Queensland is from the rest of the Australia on that matter.

	Also, I would appreciate a little more background on how the health care system works in Australia. This is minor, but I suggest that you create a new section called Objectives and list primary and secondary objectives separately from the Introduction. A significant portion of the Introduction focuses on the Aboriginal and Torres Strait Islander people. I was unclear whether this project's focus will be on analyzing this group in particular or is that only one of the subpopulations that the authors propose to study. I suggest clarifying this in the Abstract already. The Analysis section describes how estimation models will be done for several subpopulations. I suggest that the authors add a little more support for the reasons behind their decisions – especially regarding stratification on different time periods – in the Introduction with some existing evidence or provide theoretical support. I also suggest that authors adjust for multiple hypothesis testing, and that they report robust standard errors, clustered at the clinic or some other appropriate level to account for within cluster correlation. I would also appreciate some more detail on the counterfactual modeling that is mentioned in the Analysis section. I was a little unclear how authors plan to measure one's socio-economic status or income beyond patients having the low-income healthcare card or not. One way to do this (if a direct measure at the individual level is not available) is to proxy patient's socio-economic status by combining the administrative data with the Population Census at the lowest geographical level possible and use local socio-demographics (e.g., unemployment rate, income) to include in the regressions. The authors point out that one of the key limitations of the study, which is that data on non-admitted patients were not collected, results in the fact that (i) private allied health visits may not be captured once the cap is reached and (ii) that hospital services not covered by Medicare Benefits Schedule are not captured either. While that is true, I would also add that data may be missing for patients that never made it to the clinics in the first place – and they could be those that are disadvantaged the most. I recommend a short discussion on how - and to what extent - that may affect the study results.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Assoc Professor Georgina Chambers Institution and Country: National Perinatal Epidemiology and Statistics Unit, University of New South Wales, Australia Competing Interests: None declared

Comment: My main concern with the paper is that it doesn't describe the methods fully enough, particularly for those who aren't familiar with data linkage and the Australian datasets being used. Protocol paper can afford to go into greater details of the methods proposed. We have added significant detail to the methodology section, as suggested in the following comments made by the reviewer.
The introduction is relevant, but it may be worth giving more context for an international journal (e.g. Australian healthcare system, population of QLD, number babies, % Indigenous etc)

Response: A paragraph outlining the Australian health system, specifically with regards to Medicare has been added to the Introduction; demographic characteristics of Queensland have been added to page 6 in the methods section.
Suggest not splitting Primary and secondary aims The primary and secondary aims are now presented together.

Data I would like to see a diagram here of the linkages and the years, plus the non-linked survey data. I think this would be very valuable, and make the paper more interesting and easy to follow.

For each data set provide a link for people if they want more information I think more detail is needed on how the linkage will actually work, and in what environment it will be analysed (e.g. MBS and PBS have strict requirements). I would also like a formal description of each dataset, and who is Data Custodian More details on HHS funding and Costing Unit Data needed.

Additional details have been added, as suggested. A diagram showing the data to be linked has been added to page 11; we have not provided web-links as suggested due to the likelihood of web addresses changing over time, however we have added additional details; additional details on the linkage process has been added on page 10; a section outlining data access requirements has been added to page 12; a formal description of the dataset, the data custodian and the requested variables has been added to pages 6– 10.

Comment: For the AHS, what does 'similar demographic characterises' mean. Are these for women giving birth?

Response: Due to a slight change in study design, over the counter medication use has been removed, and so reference to the AHS has been deleted

Response: I am confused about the time frames for the data linkage and what data will be provided. Mothers and Babies will be identified from the PDC for 2012-2015, but outcomes will be collected to 5 years – how is this possible?

Response: Apologies for the inconsistencies in date specifications, these were typos left over from previous iterations of the protocol. The date ranges have been corrected throughout the manuscript and clarified in figure 2 on page 12.

Comment: Also, pregnancy data is also being collected, but I'm not sure how this will occur either, more details needed. Hence the need for a diagram. Furthermore, please check how you will manage the >5 years of Medicare data needed (or are your censoring mother and babies).

Response: Full details of the variables to be extracted have now been provided. Only four years of data will be extracted, from 01/07/2012 – 30/06/2016, the 'five years' statement was a typo, as detailed in the response to the previous comment.

Comment: It an overstatement that you will obtain a 'full range of health system costs' (as you later point out). Outpatient costs are missing also.

Response: This phrase has been amended to reflect the data that will be collected.
What 'clinical outcome measures' will be captured (line 36)? Administrative data isn't very good for this, so examples would be useful
The clinical measures that will be captured are now detailed on pages 6-8.
Analysis Please explain, why linear regression is being used to stratify by different SES and Geo regions. I did not understand this. This paragraph has been re-written to make the regression modelling methodology clearer.

Comment: The last 6 paragraphs of Analysis are replications with only one word change – suggestion condensing.

Response: These paragraphs have been replaced with a summary paragraph, as suggested.

Comment: Because this is a cost model, what will be used for discounting?

Response: A discount rate will not be applied, but all prices will be presented in 2016 Australian dollars, adjusted by the health price index. This is now detailed on page 12.

Limitations: What does 'once caps reached' mean?

Response: This phrase has been removed as a part of manuscript editing.

Comment: What about cross jurisdiction flows.

Response: The authors were not certain what the reviewer meant in this comment.

Reviewer 2: Sanjay K Mohanty, International Institute for Population Sciences (IIPS), India
Competing.

Comment: I am not sure whether it is appropriate to publish a study design that will be executed in later stage. Ideally, results/design are presented in scientific journal. Review is adequate and conceptualisation is strong. May be good to present a systematic review on health disparities in Australia.

Response: The paper is being submitted as a study protocol.

Reviewer 3: Tadeja Gracner, Department of Economics, University of California, Berkeley

Comment: In the Introduction, the authors present extensive evidence on how limited access to care may impact health outcomes. However, there is little information on the current status of out-of pocket costs overall and by different socio-economic groups. I suggest that the authors spend some time describing statistics on socio-economic disparities in access to care and related health care costs disparities (even if unrelated to perinatal care) in both, Australia and Queensland. That way one can get a sense on severity of the problem, as well as how different or not Queensland is from the rest of the Australia on that matter.

Response: Additional details on out of pocket expenses have been added to the introduction.
Also, I would appreciate a little more background on how the health care system works in Australia. This information has been added to the introduction, see page 4.

Comment; This is minor, but I suggest that you create a new section called Objectives and list primary and secondary objectives separately from the Introduction.

Response: A new section called 'Objectives' has been added to page 5, as recommended.

Comment: A significant portion of the Introduction focuses on the Aboriginal and Torres Strait Islander people. I was unclear whether this project's focus will be on analyzing this group in particular or is that only one of the subpopulations that the authors propose to study. I suggest clarifying this in the Abstract already.

Response: It has been clarified in the Abstract that Indigenous women and children will be just one sub group that will be analysed. Additional details on other areas of inequality have been added to the introduction.

Comment: The Analysis section describes how estimation models will be done for several subpopulations. I suggest that the authors add a little more support for the reasons behind their decisions – especially regarding stratification on different time periods – in the Introduction with some existing evidence or provide theoretical support.

Response: Additional details on other areas of inequality have been added to the introduction to support these subgroup analyses. The time groups are simply annual groupings.

Comment: I also suggest that authors adjust for multiple hypothesis testing, and that they report robust standard errors, clustered at the clinic or some other appropriate level to account for within cluster correlation. I would also appreciate some more detail on the counterfactual modeling that is mentioned in the Analysis section.

Response: An additional paragraph on the counterfactual analysis has been added to page 15. All mothers and babies across Queensland have been included in the sample, with no cluster sampling undertaken.

Comment: I was a little unclear how authors plan to measure one's socio-economic status or income beyond patients having the low-income healthcare card or not. One way to do this (if a direct measure at the individual level is not available) is to proxy patient's socio-economic status by combining the administrative data with the Population Census at the lowest geographical level possible and use local socio-demographics (e.g., unemployment rate, income) to include in the regressions.

Response: We will also identify socioeconomic status based on an area based measure, as suggested by the reviewer. Further details have been added to page 14.

Comment: The authors point out that one of the key limitations of the study, which is that data on nonadmitted patients were not collected, results in the fact that (i) private allied health visits may not be captured once the cap is reached and (ii) that hospital services not covered by Medicare Benefits Schedule are not captured either. While that is true, I would also add that data may be missing for patients that never made it to the clinics in the first place – and they could be those that are disadvantaged the most. I recommend a short discussion on how - and to what extent - that may affect the study results.

Response: The counterfactual analysis aims to, in part, account for this unmet demand raised by the reviewer. This is now made clearer on page 15 with the more detailed description of the counterfactual methodology.

VERSION 2 – REVIEW

REVIEWER	Tadeja Gracner RAND Corporation, United States
REVIEW RETURNED	08-Oct-2017

GENERAL COMMENTS	While the authors added much information on Australian's health care, I suggest that they make it more concise, as it is quite lengthy. Also, authors should link the listing of general statistics on health and socio-economic disparities with the evidence-based discussion on how access to prenatal care may mitigate them. The data section is hard to read. I strongly suggest creating a table in which key variables (e.g., as outcomes, regressors, linkage variables) and their sources are identified. Several other variables should then be listed in the Appendix. Look into combining this with the current Table 1. In the Analysis section, authors list many outcomes of interest but do not describe the hypotheses that they plan to test. I understand that they plan to test many, so I suggest that authors adjust for multiple hypothesis testing and that they mention how they plan to report robust standard errors, clustered at the clinic or some other appropriate level to account for within-cluster correlation. I would also appreciate some more detail on the counterfactual modeling that is mentioned in the Analysis section – I am still unclear how authors will execute it. The authors point out that one of the key limitations of the study is that data on non-admitted patients were not collected. Authors should add that data may be missing for patients that never made it to the clinics in the first place, and that they could be those that are disadvantaged the most. I recommend a short discussion on how - and to what extent – they anticipate this may affect the study results.

VERSION 2 – AUTHOR RESPONSE

Reviewer 3: Tadeja Gracner, RAND corporation, United states

Request: While the authors added much information on Australian's health care, I suggest that they make it more concise, as it is quite lengthy. We have only added 1 paragraph relating to the Australian healthcare system.
Response: Given that this is a model of health system use, the authors believe all information presented is essential to understanding the project.

Request: Authors should link the listing of general statistics on health and socio-economic disparities with the evidence-based discussion on how access to prenatal care may mitigate them.
Response: Statement added to page 3-4 linking the disparities in access to maternity services with disparities in health outcomes experienced by subpopulations groups in Australia

Request: In the Analysis section, authors list many outcomes of interest but do not describe the hypotheses that they plan to test. I understand that they plan to test many, so I suggest that authors adjust for multiple hypothesis testing and that they mention how they plan to report robust standard errors, clustered at the clinic or some other appropriate level to account for within-cluster correlation.
Response: - We have clearly stated the objectives of the study on page 5, which covers the breadth of the analysis that is planned from this model – there is no single specific hypothesis that is to be tested. These objects relate to discrete studies, and there will be no comparisons made between them, and as such there is no need for correction for multiple hypothesis testing.
- There is no mention of analysis that would be effected by intraclass correlation (i.e. no selection of sub-populations based upon clinics or treatment centres), and so there would be no reason to plan to adjust for such correlation.

Request: I would also appreciate some more detail on the counterfactual modeling that is mentioned in the Analysis section – I am still unclear how authors will execute it.
Response: An illustrative example has been added to page 13.

Request: The authors point out that one of the key limitations of the study is that data on non-admitted patients were not collected. Authors should add that data may be missing for patients that never made it to the clinics in the first place, and that they could be those that are disadvantaged the most. I recommend a short discussion on how - and to what extent – they anticipate this may affect the study results.
Response: As we utilised the Perinatal Data Collection to form our sample – the administrative dataset containing details of all births – rather than a service-based dataset we include in our sample all patients who may not have attended certain clinics. This is a key strength of the study, and has now been noted on page 13.