**Reviewer's report**

**Title:** The impact of the health care workforce on under-five mortality in rural China  
**Version:** 1  
**Date:** 05 Dec 2018  
**Reviewer:** Jenny Liu

**Reviewer's report:**

The authors have responded to each of the reviewers' critiques from their first submission. While the analytic rigor of the paper is improved, some of the additional details provided in the revised manuscript foster new questions about the integrity of the theoretical foundations that should motivate the analytic approach. Further, the discussion of the specific results on health professional density and their modifiers is disconnected from the discussion of policy reforms, which limits the reader's ability to actually understand what the policy lessons learned are that the paper claims to have contributed to. In many instances, the explanations provided in the Response to reviewers is better than the content in the main text. I would encourage the authors to incorporate these details and explanations in the next revision of the main manuscript. Substantive areas for improvement are listed as follows:

* **P6 7-11:** Can you provide more information about how these contextual variables were used in the studies you cite? What were the findings regarding these factors? Are they moderators or just control variables? Is this somehow connected to your claim in the next paragraph regarding studies using fixed contextual characteristics?

* **P6 36-41:** You argue the pace of economic growth is what actually matters for U5MR. Why do you say this? While this may be a gap in the literature, there may be good theoretical reason why level of economic size is used instead of the rate of growth. Does theory support your assertion that growth matters? Please say what that is to justify your hypothesis and motivate your analysis from theory.

* **P6 49-60:** Why only rural China? What period are you referring to (2009 to what)? In general, more information on China's health reforms in needed to motivate your paper and your hypothesis in the introduction section. The authors added this background information in the discussion, but this should be moved to the beginning of the paper to motivate and understand how your conceptual model relates to your empirical analysis for the specific context in which you are assessing. The background may also help to support your choice of variables and argument for why some factors are theoretically important for your empirical analysis and why others do not matter as much.
The clarification of medical education for physicians and associated reason for which categories are included in the Response to reviewers should be included as a footnote in the main text where appropriate.

P7 39-54: "8137 counties." These are county-year units. Please be precise.

What percentage of the observations were imputed or excluded due to being outliers? Please put in the main text.

Yearly average population for all counties = 894M? For average population for each county each year? Please clarify.

Fig 1:

What do the dashed boxes represent? Please explain this figure in words in the text.

Please also caveat that the variables you have listed in the boxes are only illustrative - these are certainly not comprehensive as your writing seems to suggest (e.g., "all identified confounders), and it is unclear why these specific factors are included in the diagram as opposed to others (is this motivated by your literature review)? Your Response to reviewers has a better explanation, but still does not thoroughly describe how you developed this conceptual model. Please explain how this diagram is motivated.

It seems that most factors listed in the diagram are high-level concepts (e.g., health professionals) whereas those listed as "confounders" (i.e., female illiteracy, skills mix) are more specific. The level of abstractness is not consistent within your diagram.

I disagree with your assessment of the role of skills mix (as represented by your ratio of physicians to nurses) in the causal pathway to U5M. Skills mix is an essential part of the production function for health services, potentially as equally important as the number of professionals. While skills mix may not be the main variable of interest in our analysis, your description of the importance of this variable is inadequate. It is one aspect of "health professionals," rather than simply an afterthought and confounding variable.

Your explanation of time-varying confounders in your Response to reviewers should be included in the main text. These details are essential to evaluating the integrity of your empirical specification.

Female illiteracy rate:

Clarify definition. Is this the % of females over age 15 who are illiterate?

Table A1 indicates that there are only 71 missing observations for this variable, but Appendix A3 says that only data for 2010 is available. These two statements seem contradictory. Please clarify.
o If only data for 2010 is missing, then does it make sense to include the variable at all given the amount of missingness and the need to impute nearly all of the data? What happens to your estimates when this variable is excluded?

o Theoretically, how important is this variable. What does it represent?

* Why are some variables logged and others used in levels in your main specification? Please give rationale, such as that given in your Response to reviewers for density.

* There is little discussion of how your empirical specification relates to your conceptual model in Figure 1. Here you can also explain why the inclusion of county fixed effects is necessary to control for other factors that you don't have measures for.

* Thank you for including additional sensitivity analyses. However, the following issues arise in your presentation of these analyses:

o You discuss including sanitation and hygiene infrastructure and skills mix in your sensitivity analyses, but you don't say how you used these variables in your empirical specification in the main text. Please explain. It is not enough to say they are important; please say why are they important.

o What happens if the density of nurses and physicians are included as separate variables rather than as a ratio?

o What about the density of other nurses and physicians to other types of health professions? It is unclear why you chose the particular ratio of physicians to nurses as opposed to other types of cadres. What is the variation across counties and over time in health worker availability by cadre? Figure A1 does not completely address the skills mix distribution.

o Please add the extended discussion of the results of your sensitivity analyses contained in the Response to reviewers into the main text.

* Do registered health professions translate directly to health professions that are actually working in patient services? This measurement detail was not addressed.

* P13 30-51: More information is needed for the explaining the differential findings by poverty and economic growth.

o Thanks for providing more background information on the 2009 reforms. Some of this basic information should be moved to the introduction to ground your theoretical and empirical approaches.

o While you explain some of the specific policy and programmatic changes that occurred with the 2009 reforms, please explain how they were targeted? Were some purposefully designed
for poverty or slower growth areas? Why was that done? Or why did these reforms differentially affect poverty areas?

o You say that more physicians were recruited to poor areas. Why only physicians? What about other cadres? What does that say about your main result for your health professionals variable? Are your results driven all by the density of physicians? Certainly, some of the primary prevention services you list likely do not require the skill level of physicians to carry out.

* Please revise the text for a more informed and nuanced discussion of what your results actually mean in light of the historical reform context and in light of the variations in these reforms across counties (since this level of variation is what you argue is actually important and a main contribution of your paper). These details are critically important for extracting the concrete policy lessons learned for other developing countries. In the conclusion, you say that the results should be instructive for other countries, but you don't say exactly how your results should be instructive. What is the most important takeaway points for policy? Can you say something specific about density, reforms, targeting, and/or county-level variation?

A close editing of the paper is needed. Some places require tempering of strong language used and more precise interpretation and critical description of analyses. Other places need clarifications. Some examples follow, but there are many more instances where editing attention is needed:

* P5 26: "...human resources..." for health?


* P2 6: "density of health workers/doctors)..." Clarify: is this both health workers and doctors? Workers per doctor?

* P6 20: "national or state/province level, including China." Refers to studies on China? Using data from China? Clarify. Same sentence structure is seen at P6 47.

* P6 22: "impact of health care workforce..." Impact is not in fact what these studies were estimating. There are many other instances of this throughout the paper.

* P6 23: "...limiting the inferences that can be drawn from such studies..." This perspective depends on what the hypothesis is. Perhaps you are trying to say that this paper's analysis limits to inference that you are seeking (and not necessarily the inferences that the paper's authors were seeking).
Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Quality of written English
Please indicate the quality of language in the manuscript:

Not suitable for publication unless extensively edited

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal