Reviewer’s report

Title: Evaluation of stability of directly standardised rates for sparse data using simulation methods

Version: 0 Date: 22 Mar 2018

Reviewer: Aaron Osgood-Zimmerman

Reviewer's report:

Introduction:

The reviewed manuscript studies the accuracy of directly standardized rates (DSRs) under a few varying conditions with particular focus on the minimum number of observed events required to confidently generate accurate DSRs. The authors simulate observed event data over 19 age groups assuming independent Poisson distributions under different scenarios where the total number of events, age specific risks, and the population distribution across age groups were changed. For each experiment, they perform 10,000 simulations and calculate the DSRs and confidence intervals (CIs) using three different approximations to construct CIs. They then assess the accuracy of both the DSR estimates and the different CI constructions to study under what conditions DSRs can be trusted and which CI calculations are most robust.

Merits:

The authors appear to have identified a relevant real-world problem: DSR estimates are often published for rare events (or more generally for events with low recorded counts) and it is unclear how trustworthy these estimates may be. The authors aim to build upon prior work by Ng et al [1] who performed a similar simulation study but in [1] they did not simulate and study the accuracy of DSRs for scenarios with fewer than 10 events occurring over the 19 age groups. This manuscript looks at the sensitivity and accuracy of DSR estimates and associated CIs under a range of conditions but in particular they aim to add to the discussion around best practices for estimation and interpretation of DSRs in these very small event count scenarios. This seems to me to be worthy study goal - although not particularly novel considering that it is mostly a minor extension to an already existing study - with results that would be of interest to the wider community.
Critique:

I felt the motivation of the entire paper to be quite lacking. Although based on earlier work, this manuscript should still be able to stand alone for the interested reader.

I believe that more explanation, motivation, citations, and/or justification is needed for:

- when and how DSRs are used.
- when in practice the small event counts may be of concern and how pervasive an issue this may be. Considering that this paper is primarily focused on this aspect this motivation seems particularly lacking.
- the definition of 'accurate', 'conservative', and 'liberal' CIs. Either a citation is required or the authors need to make it more explicit that these are definitions they are defining specifically for this study.
- the different formulas used for CIs. [1] goes into a large amount of detail discussing the motivations between different methods for estimating CIs including some explanation of their derivation, assumptions that are made, and the issues that a particular CI estimate aim to address. This sort of description both acknowledges relevant prior work and provides the reader with a grounding to understand and interpret results from the simulation study.
- why only a (small) subset of the CI formulas used in [1] are used in the study.
- the settings used in the simulation study.
- the authors' suggestion that 10 or more events is the suggested minimum number of events for producing DSRs. They state that: "This means that if 10 or more events are observed the expected number of events is extremely likely to be below 5. Therefore if the observed number of events is 10 or more the 95% CI is very likely to be 'accurate'." While this appears to be a safe decision, why have they picked 10? Is there a slightly lower threshold that would be acceptable so that more DSRs can be produced? Can the authors more precisely quantify through their simulation what this statement means? Given that this is the crux of the manuscripts conclusion more work needs to be done to justify the "recommendation that at least 10 events must have occurred for a directly standardized rate to be published."

Along these lines, some of this information is scattered throughout the paper (e.g. the second paragraph of the results section briefly touches upon some practical differences between the
Dobson and Tiwari methods) in unusual places. Consolidating these concepts to a prior work or methods section instead of interweaving them into the results and discussion would make for a much more coherent read.

Since this paper is mostly an extension of [1], I would like to see comparisons of the results between the two. Are the same conclusions found in [1] found in the new simulation studies under the different simulation scenarios?

Conclusion:

While I believe that the authors have set out to answer a useful question, I don't believe that the work, as it stands, demonstrably, clearly, and effectively answers the question: what is the minimum number of events required to safely publish a DSR. This is especially true given that the work is an extension of existing studies which already aim to address which CI constructions are robust in different settings.

I believe that the authors could fortify the existing work into a useful publication if they take the time to reorganize, motivate, and solidify the results. That said, I think there are a number of interesting scenarios which, to the best of my knowledge have not yet been studied in a simulation setting, which could extend this work into a novel study instead of a (relatively small) extension of existing work.

One particularly interesting expansion would be to study the independent Poisson assumption. The authors claim that "for most diseases and causes of death this is likely to be a reasonable assumption." While there may be cases where this is reasonable, I would disagree that it is `likely` to be the case. Many non-infectious diseases have causes that are related to spatial covariates, and as such all individuals, from all ages, within the same spatial unit (e.g. district) will not be independent if the relevant spatial effects are not taken into account. I think it would be quite interesting for the authors to simulate spatially correlated data (e.g. using a Besag-York-Mollie type model) and then to run their assessment framework, assuming independence, on the non-independent data. They could then study under what sorts of correlation settings the DSR and CI estimates are reasonable.

Another suggestion would be for the authors to expand upon what happens when fewer than 10 events are observed. Are the results simply not published? Can data from multiple neighboring districts be aggregated together until the minimum event count is reached? What repercussions does this cause for the validity of the estimates?
References:


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:

Needs some language corrections before being published

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