Reviewer's report

Title: Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods

Version: 1 Date: 10 December 2013

Reviewer: Michael Batz

Reviewer's report:

This is a very good review article that delves into an important subject that isn't often described in the literature in such detail. With some modest clarifications and changes, I think it will be an excellent addition to the literature.

DISCRETIONARY REVISIONS

1. The only overarching issue I had was the lack of discussion about uncertainty, either in general or specific to the studies themselves. It is not just that surveillance data under-represent all illnesses, it's that quite often we don't know how badly they do so. It's difficult to know what you don't know, which is why there are so many different approaches that are being employed to get at underascertainment and underreporting. The authors know this of course, but it doesn't quite come through in the text.

Moreover, on the point of uncertainty, I think the authors might consider discussing that these many factors are multiplicative, which is why they explode so rapidly, and why even small degrees of uncertainty in measuring individual components of underestimation can lead to very wide ranges in MF when they are all put together to get incidence estimates. This is something that also doesn't quite come through in the text.

There is also very little uncertainty reflected in the tables. I am fairly certain that a number of these studies provide not just expected values for underreporting or underascertainment or both, but do provide either ranges or variance. I think addition of known information on uncertainty around MF estimates would be very helpful.

2. p.16 - Re: mortality - I would add that in some countries without nationalized health care, medical data can be biased by financial incentives. For example, in the United States, reimbursement by private insurance or federal Medicare/Medicaid programs may be related to the ICD code used for primary diagnosis. This may not be an issue in Europe, but there may be similar internal or bureaucratic reasons for mis-coding or massaging reporting data on cause of death.

3. On the importance of sex- and age-specific multipliers: The text points out ways in which both underascertainment and underreporting can vary by age- and sex-, and in numerous places suggests that the best UE studies account for both.
However, they only turn up one study that accounts for age, and mention none that account for gender. It is therefore striking that the authors do not make a bigger deal out of this gap in the literature in the discussion and conclusions sections, and I strongly suggest adding a bit more to bolster this area.

MINOR ESSENTIAL REVISIONS

4. p.3 - Background, line 4 -- I think "limitation" might be a more accurate word than "restriction."

5. p.13 - "ascetained"

6. p.14 - Recommend changing "discussed" to "discuss"

MAJOR COMPULSORY REVISIONS

7. p.4 - Definitions -- This whole section could be strengthened a bit. Since the section is titled as definitions, I feel it should provide complete, standalone definitions, and not just descriptions. I also feel the definitions provided could use some work to make them more precise.

8. P.4 - Underestimation -- Underestimation is defined as "an umbrella term encompassing all forms of incomplete data collection at all levels of the surveillance system" but I find this awkward in a couple of ways. First, defining a word as "an umbrella term" seems circular -- you've said that a word is a word - and doesn't help to clarify what exactly it is. Is it a phenomena? A principle? Moreover, I don't think it's fair to say that underestimation reflects incomplete data collection. This suggests that it is somehow a failure of surveillance systems to not capture every infection, but this is preposterous. A surveillance system aimed at collecting hospital data is NEVER going to capture asymptomatic cases or symptomatic cases who self-medicate or ride it out without seeking care. Isn't underestimation about interpretation? That is, the data is what it is, and it is how one uses it that defines whether it needs to be adjusted. If you are using hospital surveillance data to talk about hospital utilization, there is no underestimation because you aren't extrapolating to the broad population. All told, I'd probably rewrite the first definition to be something more like "Underestimation (UE), as defined here, is an umbrella term that can be understood as the many ways in which surveillance systems fail to reflect all infections in a given population; mathematically, it is the number of infections estimated to have occurred in some population over some time period for every known illness or infection captured by surveillance in that same time period."

9. I wonder whether it might be worth defining "surveillance system" and "notification system." In my own parlance, the surveillance system starts at the point of reporting -- that is, a doctor failing to diagnose an illness due to lack of taking a sample, lack of requesting the right test, or due to lab method sensitivity/specificity is hardly a failure of surveillance. That is a failure of treatment, or of health care, or whatever. It becomes a surveillance "failure" only when the diagnosed disease does not translate into a record within a data system that is tracking such things. Put another way, I see underreporting being
separable into two causes -- the failure of the health system to diagnose, and the failure of the surveillance system to capture all positive diagnoses. The text lists all of these causes of underreporting but kind of lumps them all together as one, which to me suggests all of these issues are the cause of surveillance systems when a good many of them are not. Perhaps the issue is using "surveillance" and "surveillance system" too loosely or interchangeably, sometimes meaning the infrastructure/database and sometimes meaning the bigger idea of counting illnesses.

10. p4. Underascertainment: Similarly, the sentence that begins "UA comprises of non-healthcare seeking infected individuals" is awkward for a couple of reasons. First, 'comprises' is used to describe parts of a whole, and individuals are not the individual components of a whole underascertainment.. Second, grammatically it should be "comprises" and not "comprises of" (e.g. "a full pack comprises 52 cards.").

11. p.15 - Paragraph beginning "However, adjusting for underestimation..." -- I found this paragraph to be strangely out of place. Perhaps it is how I read it, but I found it somewhat offensive. I don't feel that discussion of "political" worries has much place in a scientific journal and I disagree wholly with the assertion that discussion of a scientific principle, such as public health underreporting and underascertainment, is an implicit criticism of datasets or surveillance systems. The last sentence is upsetting because it reads like the authors are apologizing for doing their work -- it puts far too much responsibility for misinterpretation or misunderstanding on the shoulders of those producing estimates, because I promise you no matter how good of a job the analyst does there is going to be someone out there who doesn't like it or doesn't understand it or whatever. Of course, trying to communicate this work is important and there may be reasons for concern. But let's not bend over backwards here. I recommend boiling this paragraph into something more conservative, like "Because adjusting for underestimation results in higher disease burden estimates and can result in differing rank orders of public health importance compared with unadjusted surveillance data, care should be taken in such studies to clearly communicate both the need for such adjustment and the methodologies employed."

12. In my opinion, Tables 1 and 2 are the heart of this article. As such, I think it's a significant missed opportunity not to list study type for each of the studies listed in these tables. It would require additional rows for countries with multiple studies, as are done in Table 3, but it would be invaluable information, particularly since so much of the text is dedicated to describing these various methods. I don't think it would be too difficult to add some columns, particularly if some of the text-heavy pieces were moved to footnotes, and if abbreviations were used in the table and expanded in a note. Plus, it's hard to read with the references right next to the numbers. Having them in another column would help readability, and would allow for the inclusion of more uncertainty ranges, when available.

13. Table 3: I don't really get much out of this, particularly if Tables 1 and 2 were cleaned up and expanded to include study types. I think it might be worth
exploring some of these values in a figure that could highlight some of the differences. Perhaps I'd get more value out of it if the text on p.14 was a bit more detailed, but as is, what is the story other than breadth? And how do I not also get this out of Table 1 & 2? Another aspect that could be interesting would be to show how much MFs differ across countries, perhaps simply showing the results from reference 74, Havelaar et al. 2013, across EU nations.

14. The figures are largely duplicative, with a lot of them highlighting the same ideas and framing the issues only slightly differently. Figure 2, in particular, seems without much justification at all. Figure 3 could be useful if the text described the variables that are shown in the diagram.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.