Reviewer's report

Title: Internal validity of the Swedish Maternal Health Care Register

Version: 1 Date: 2 January 2014

Reviewer: Russell Kirby

Reviewer's report:

In this manuscript the authors provide results of a data quality analysis of the Swedish maternal health care register. The manuscript is generally well written, but could be improved in several areas.

Most notably, the results are presented in terms of percent agreement and correlation coefficients - most of the similar literature in this area uses more sophisticated measures, including kappa, sensitivity, specificity and false positive rate. Kappa is mentioned as a measure in the methods, but no kappa statistics are provided in the results or tables. However, the authors in several places refer to 'degree of coverage', a non-conventional term for research in this area. Thus it is difficult for readers to compare these findings with those of other studies. The authors are correct that, based on their presentation of the results the data do seem generally of sufficient quality for use in quality improvement and perinatal research studies.

Second, the study is not well integrated with the broader literature. To make a contribution internationally, the authors should couch their study in this broader context, referencing appropriate high-level contributions, and comparing their results to specific studies that have examined data quality for similar variables. An example of a similar study (though for a single hospital only) in Costakos et al Am J Perinatol (1998). A recent paper in Medical Care by Lain et al (2012). Two commentaries examining broader issues are Kirby Am J Epidmiol (2001) and Ananth Am J Obstet Gynecol (2005).

Third, the authors could be a bit more specific in their discussion of the study design. On p 8 top, there is mention of a power analysis, but no details how it was done. On what basis was the sample deemed sufficiently large? It probably is for relatively common outcomes, but an n < 1000 probably isn't large enough to detect more infrequent yet important outcomes of pregnancy and delivery care. Also, the selection of hospitals hardly provides a representative sample of the population so this should be highlighted as a potential weakness. The inclusion criteria that data on the study subject had to be in both sources is of course essential to conduct the data validation study, but . . . please inform the reader how it comes to pass that 15% of pregnancies in the study year were not included in the register (p 9)? This is very concerning given that the MHCR should by definition include all deliveries? Are there any systematic biases related to this that readers should know about?
A few other things to consider:

What is the definition of a ‘quality register’? (p 5 and elsewhere) Readers will mostly be familiar with the concept of a registry, but what is the difference between say the Swedish Medical Birth Register and a ‘quality register’ for perinatal outcomes?

Re prenatal care, some years ago this reviewer provides a comprehensive critique concerning collection of data on trimester when care began and number of prenatal visits (Kirby, Paediatr Perinat Epidemiol 1997), the authors might wish to consult this in thinking about factors involved in the discrepant results.

A more systematic discussion of sources of error might be useful (p17 -18), as some of the issues involve transcription and manual data entry while others may reflect more serious issues (but either way the end result is discordant data across the two sources).

Table 1 demonstrates the issue that sample size needed varies depending on the frequency of the item of interest. All women have a country of birth, but only a small proportion were using snuff at first prenatal visit. Likewise, a small proportion had an amniocentesis.

In Table 2, the last column shows the percent of the total sample (N=878) from each hospital. Very nice to know that each contributed 10-11 percent, but that is expected. Readers might be more interested in the proportion of annual births at each hospital included in the sample, which seems to range from less than 1 percent to almost 8 percent.

In Table 3 and 4, it would be nice to see the kappa results, as well as sensitivity/specificity. There are some additional items that warrant mention in results/discussion. Maternal weight at last data entry after 35 weeks - why is this only available in both sources for 86.6%? With our current interest in weight gain, obesity influences on labor and delivery, and related concerns, one would think this would always be documented . . . or perhaps some of the women delivered prior to 35 weeks in which case the indicator should only be assessed for women with a term delivery. Likewise for the 2 hr OGTT - is this because other tests are also used in clinical obstetrical practice? And why is whether the CS was emergency or elective not reported in 1 of 8 cases in the medical record?

And finally, it is unclear that the figure adds anything to the paper - All of this information is already adequately discussed in the text.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:

'I declare that I have no competing interests'