Reviewer's report

Title: Variation in Classification of Live Birth with Newborn Period Death versus Fetal Death at the Local Level May Impact Reported Infant Mortality Rate

Version: 3 Date: 18 January 2014

Reviewer: Lucy Smith

Reviewer's report:

Major revision
The authors seem to have misinterpreted my discussion of the statistical methods. I was referring to the need for them undertaking a multilevel logistic regression, and not that they should aggregate hospitals into clusters. As the analyses stand, the assumption is made that all of the births within a hospital are independent and yet this is not an appropriate assumption. Births at the same hospital are more likely to have the same characteristics than births from the population at large e.g. due to the local age distribution of women, socioeconomic deprivation of the area, ethnicity of mothers, specialisms of hospital doctors, hospital protocols etc. Multilevel logistic regression recognises these hierarchies. If a two-level logistic regression model is fitted, this would allow for grouping of birth outcomes within hospitals and would include residuals at both the birth and hospital level. Therefore the residual variance is separated into a between-hospital component (the variance of the hospital-level residuals) and a within-hospital component (the variance of the birth-level residuals). The hospital residuals, or ‘hospital effects’, represent unobserved hospital characteristics that are associated with birth outcomes. Even though the models in the paper adjust for known case-mix variables, it is the unobserved variables which lead to correlation between outcomes for births from the same hospital.

If the hierarchical nature of the data is ignored, using standard logistic regression as here, the standard errors of regression coefficients will be underestimated. Consequently this will lead to an overestimation of the statistical significance. Standard errors for the coefficients of the hospital-level variables will be the most affected by ignoring the hierarchical grouping.

The process of undertaking this analysis compared to that undertaken so far is relatively trivial and is available in standard software packages such as SAS, Stata and R. A multilevel model analysis would result in an extremely important and interesting paper.

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

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

I declare that I have no competing interests