Reviewer's report

Title: The Danish Neonatal Clinical Database is valuable for epidemiologic research in respiratory disease in preterm infants

Version: 2 Date: 30 October 2013

Reviewer: Stephen A Pearlman

Reviewer's report:

This is a well designed study to ascertain the accuracy of data entered into a national database by comparing the data originally entered by clinicians to data from the nursing records. The methodology seems appropriate and the data is sound. I think the syntax of the writing is very good. I wonder why the authors decided to just look at the veracity of the respiratory data instead of all the data in the database. Nonetheless, I think this is a worthwhile endeavor to assess the validity of their database for use in subsequent studies.

Minor Essential Revisions

1. There is no description of how the data was extracted from the nursing records until the discussion section (bottom of page 6). That must appear in the methods section (saying that it was one of the authors that did it etc.)

2. They need to tell us how the clinicians get the data to fill out the form at discharge. Are they just relying on their memory or do they use certain sources.

3. The wording at the bottom of page 5 confused me. They characterize the difference between the database and the chart extraction to be a "minor underestimation" for nCPAP and ventilation therapy. Next they tell us that this is not true for oxygen therapy yet in the next sentence they tell us that the concordance was "low" for all 3 variables. This needs clarification. One way to improve this would be to leave the word 'minor' out and just state the facts allowing the reader to draw the own conclusions about the data, I don't understand why they say the same is not true for oxygen therapy but the concordance is low.

Discretionary Revisions

Explain why you conclude that the database is valid and can be used as a research tool if the continuous variables do not show good concordance. Perhaps part of the discussion should be changing the methodology of how the data is first entered to make it more accurate. (i.e. have the clinicians refer to the nursing records.

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

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.