Reviewer's report

Title: Harmonisation of datasets structures prior to conducting statistical analyses with multiple datasets: an automated approach.

Version: 2 Date: 1 April 2011

Reviewer: Tolulope A Oyetunji

Reviewer's report:

1. No major compulsory revision

2. Minor essential revision:
   - The structured.dta data look different from what is cited on the table in the manuscript. This may be a cross over from the manuscript prior to review. Please verify.
   - Regarding the exception.dta, my understanding is that you may go ahead and use these variables (or not) depending on the reason for the exception e.g. if wrongly coded, you can then simply accept. If not existent, there is no need for search B. Please make this clear if it is so. The name of the dta file may connote exclusion.

3. Discretionary revisions: The manuscript reads better and the ability to run the datasets makes it easier to follow the manuscript. I am curious to know how this will perform with very large datasets with millions of records and maybe 10 datasets. Can the author comment on this? I know you mentioned 50 variables and 250 datasets.

Overall, a very novel approach to data harmonization. No doubt, there will be repeated need in the future to improve on the methodologies as more researchers utilize this software. This is similar to phase IV clinical trials, should not stop it from getting published.

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

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