Author's response to reviews

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

Authors:

Xavier Bosch-Capblanch (x.bosch@unibas.ch)

Version: 3 Date: 4 April 2011

Author's response to reviews: see over
Dear Adrian Aldcroft,

Please, find enclosed the revised manuscript with the title “Harmonisation of datasets structures prior to conducting statistical analyses with multiple datasets: an automated approach” for your consideration.

I have received comments from two reviewers, one whom does not include any further criticism. The remaining minor comments are addressed below and changes shown in the manuscript as track changes or comments. I have also addressed editorial’s requests.

I hope that these changes fulfil the reviewers’ and editorial expectations.

Looking forward to hearing from you, sincerely yours,

Xavier Bosch-Capblanch

<table>
<thead>
<tr>
<th>Tolulope A Oyetunji</th>
</tr>
</thead>
<tbody>
<tr>
<td>The structured.dta data look different form what is cited on the table in the manuscript. This may be a cross over from the manuscript prior to review. Please verify.</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>I am curious to know how this will perform with very large datasets with millions of records and and maybe 10 datasets. Can the author comment on this? I know you mentioned 50 variables and 250 datasets.</td>
</tr>
</tbody>
</table>