

## **Appendix 2: De-duplication Process**

Whenever two databases are compared, the following procedure is used in Excel:

1. Add a column to each database to be compared that specifies the name of the database. Ex: PubMed, Scopus, etc.
2. Combine the databases by copying and pasting
3. Sort by Title> Sort on values>sort ascending order> Then by Author> Sort on values >sort ascending order
4. Use conditional formatting to highlight duplicates
5. Delete the duplicates from the database that is compared against to find the entries that do not exist in that database. For instance, if the purpose is to find entries that do not exist in database A but exist in database B. Delete the duplicate whose column value – created in step 1- is A.
6. Sort by Title> Sort on values>sort ascending order> Then by Author> Sort on values >sort ascending order
7. Do manual check of titles and authors to see whether Excel missed any duplicates. If it did, remove the duplicate

Using the above procedure, first each non-PubMed database is compared to PubMed to find entries that are not in PubMed. Then, all these entries are combined in a single excel file. Afterwards, using the above procedure one more time duplicates are removed. In the final step, these de-duplicated entries are combined with PubMed database to obtain one grand database that includes all unique entries.