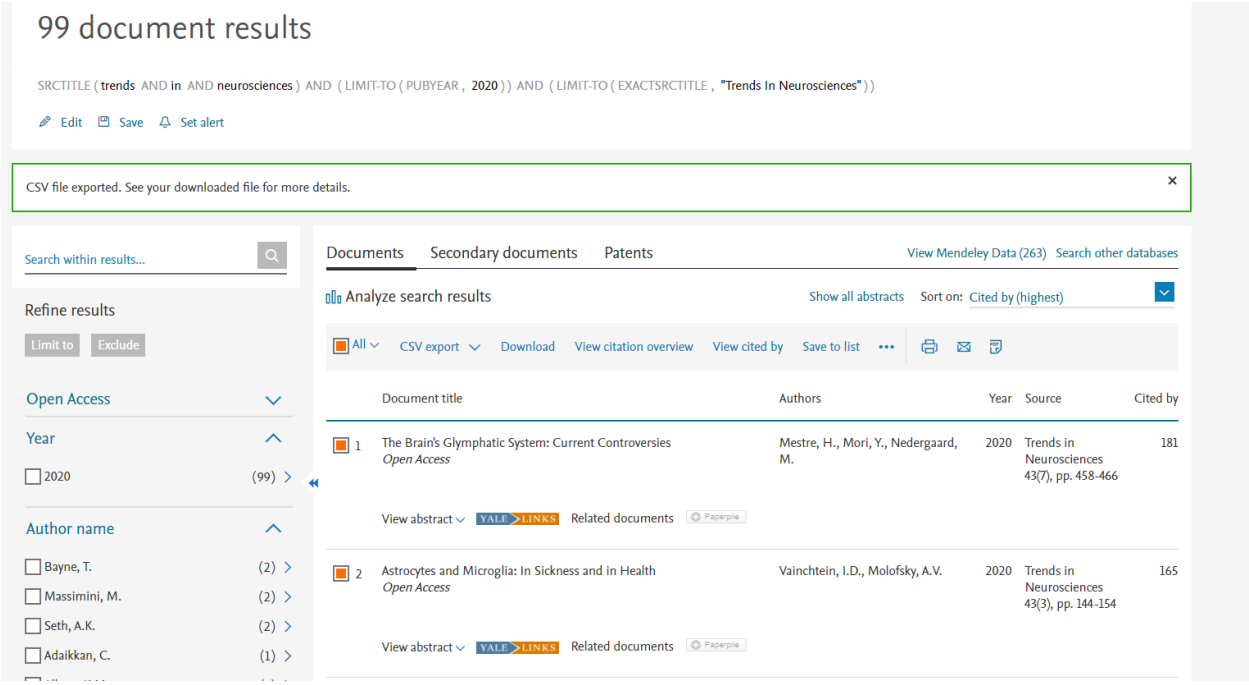
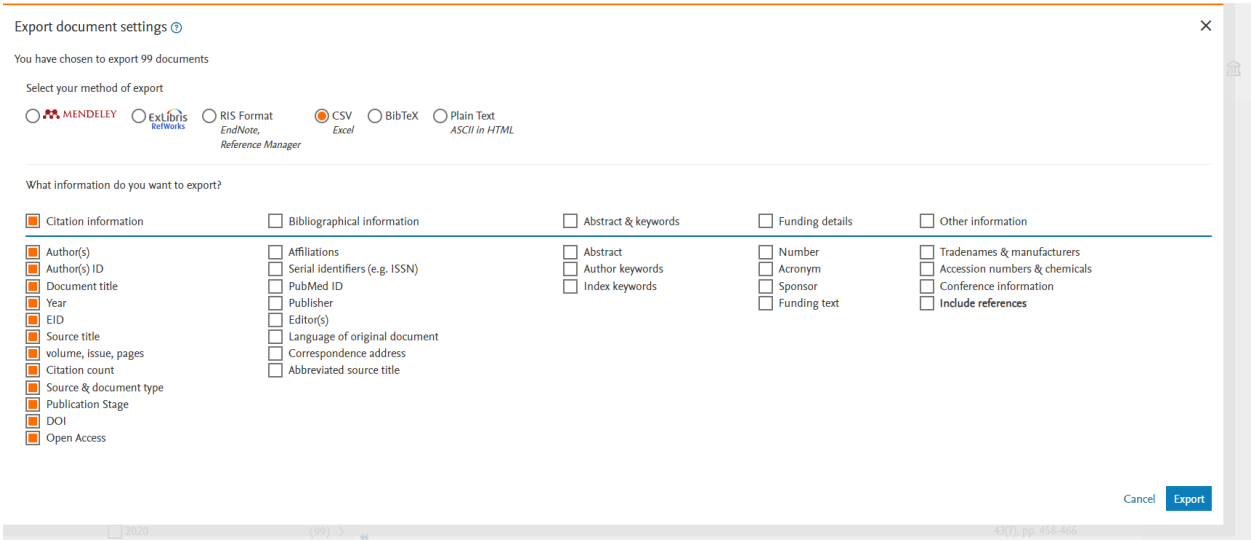
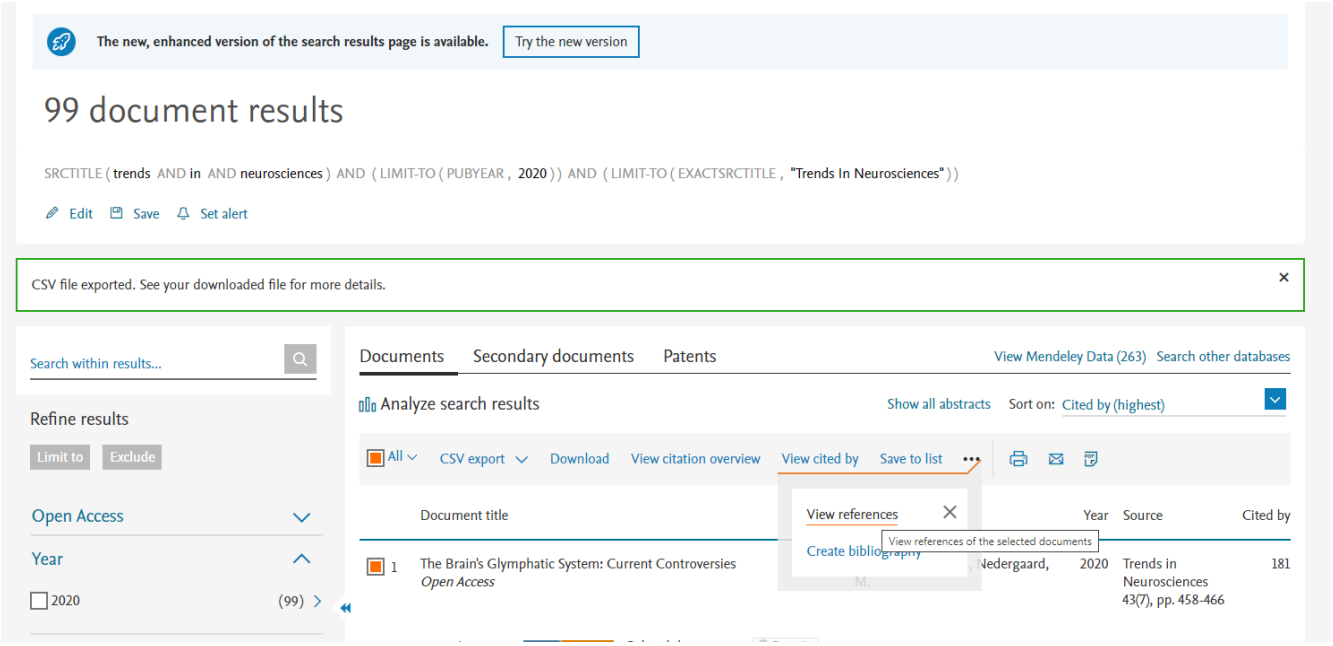
1. Go to the Scopus website and search for relevant papers
2. For example, we searched for each journal of interest by each year (example below: *Trends in Neurosciences*, 2020)



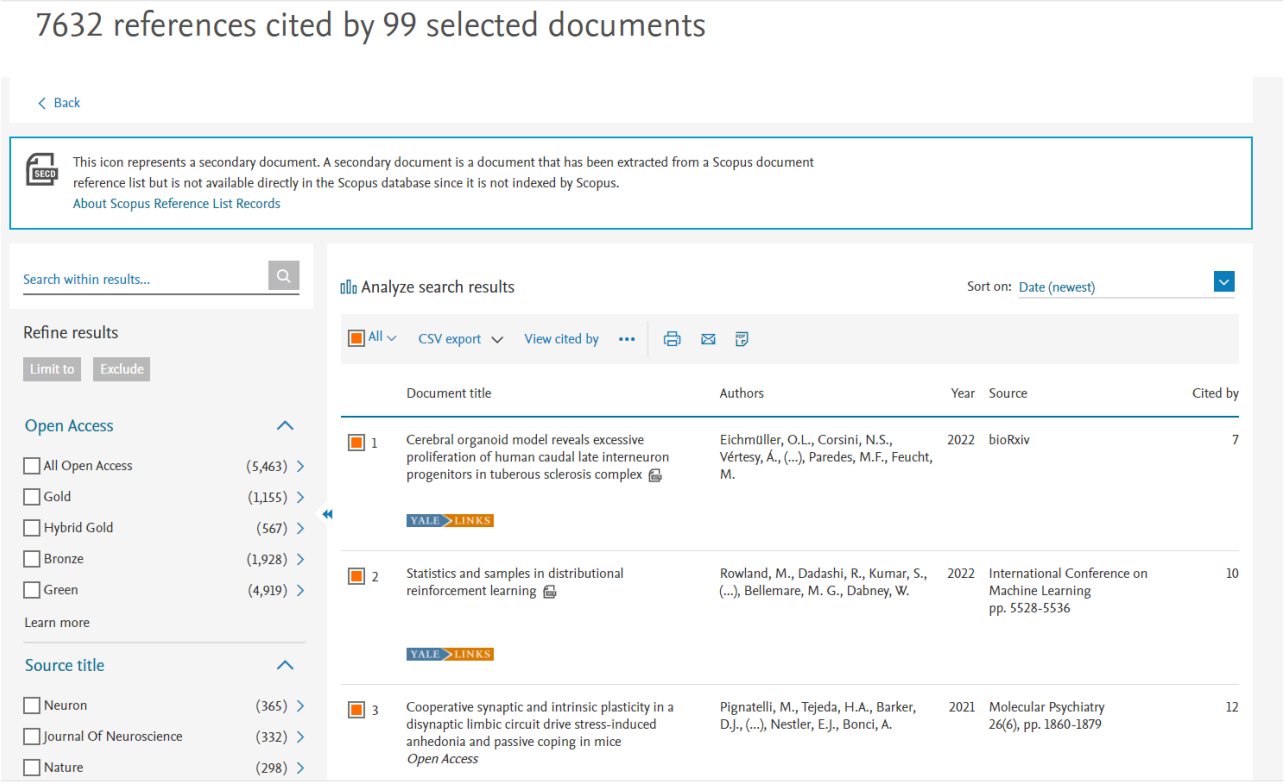
3. Click “CSV export” and export the data:



4. View all references by clicking “view references”



5. Export all the references, as done above. We follow these two steps (i.e., two separate downloads) to establish an “article database” and a “reference database.” This saves time when downloading data through the Pybliometrics API



6. Run sc_journal.py, modifying the folder structure as necessary. In our case, we had three folders (“All_Neuro”, “All_Neurology”, “All_Psychiatry”). Within each of those three

folders were folders for each journal, which included both the article database .csv files and reference database .csv files for many years.

7. Run get_auth_info.py. This will add author information to the self-citation information obtained in the previous steps.
8. Use genderize.io to probabilistically assign gender to each name. We recommend using the .csv tool, but you may also use their Python API. Please note that this requires a paid subscription.
9. Finally, perform the desired analysis. You can see our analysis in the notebook bootstrap_analysis.ipynb.