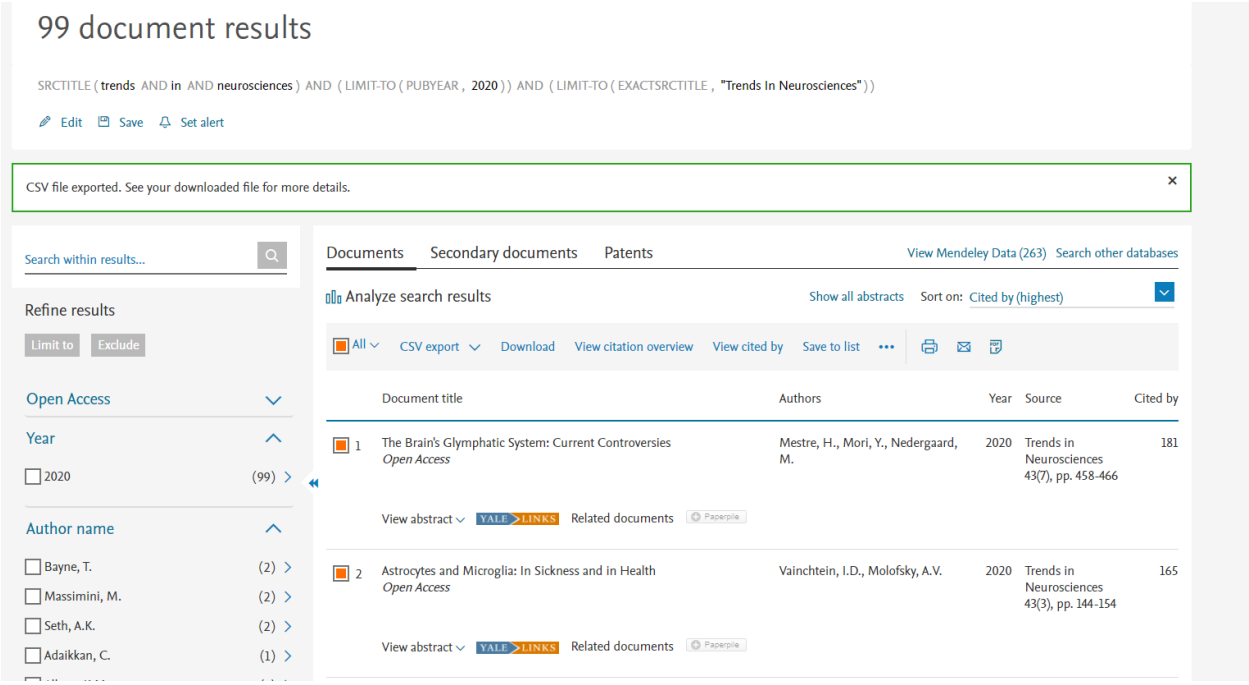
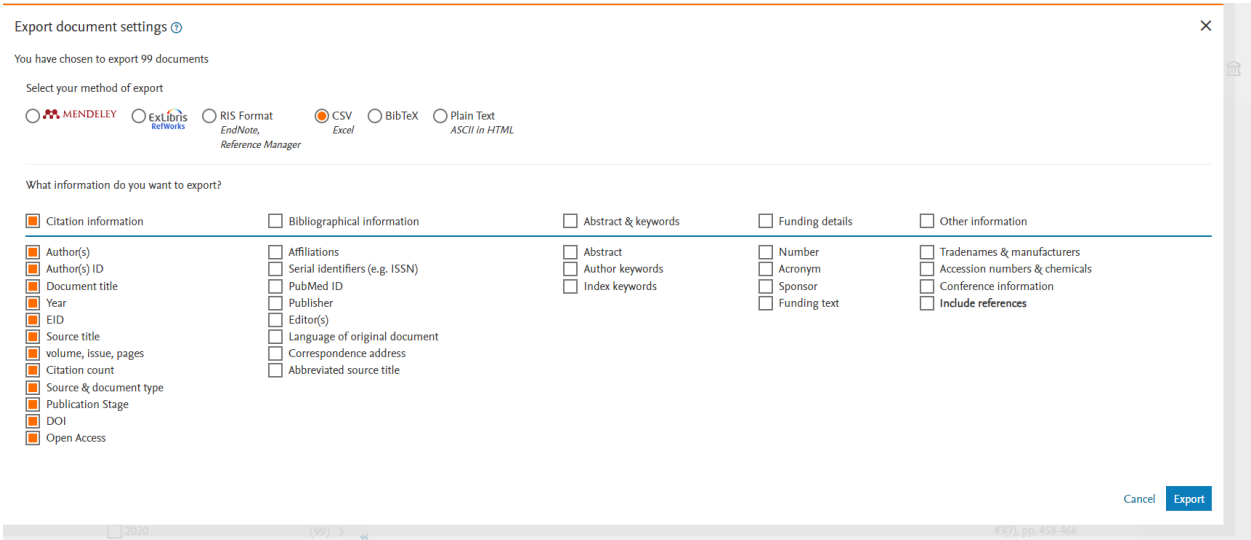
1. Go to the Scopus website and search for relevant papers
2. For example, we searched for each journal of interest by each year (example below: *Trends in Neurosciences*, 2020)



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



4. Run sc_by_pair.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 the exported .csv files. (**Note:** the download requirements differ from those in the “raw_data_analysis” folder because here the database is at the level of a citing/cited pair, not at the whole-article level).

5. 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.
6. Run model_building.py to build logistic regression models. You may enter the “--y” argument as 1) “pairs,” in which case it runs the citing/cited pairs model, or 2) “extreme,” in which case it runs the highly self-citing model