Dr. Downing asked about how our efforts in Biostatistics lead to publications. To answer his question, we need to link our BIMS project data to our list of PubMed publications. As you know, we don’t prospectively enter information about our publications into BIMS, so those links need to be retrospectively defined now. We need your help in defining those links. We’ve done most of the work to make it as fast and easy as possible for you to do your part.

We obtained bibliographic data for all 12,604 PubMed articles with at least one St. Jude author affiliation and extracted the subset of 2,854 articles that have at least one of the following three indications of authorship by someone in our department: an author name of a paper with a St. Jude institutional affiliation matches that of a department employee in BIMS, listing of an author affiliation with the St. Jude Department of Biostatistics, or listed in one of our NCBI MyBibliography pages.

We then matched the authors of those 2,854 articles with the initiator, sponsor, and requestor of 4,362 collaborative data analysis BIMS projects. We matched project personnel and title words of each of the 12,449,148 possible pairs of BIMS project and PubMed article and retained up to the 5 best BIMS project matches for each publication. These pairs are listed in the *pair.data* spreadsheet of the Excel file *BIMS-PMID-matches-2021-04-12-for-dept.xlsx*.

Some PubMed publications and BIMS projects did not have a meaningful word match score. The data for these publications and projects are provided in the spreadsheets *matched.pubs* and *unmatched.BIMS*, respectively. To help find matches for these publications and projects, the spreadsheets *all.PubMed.data* and *all.BIMS.data* provide data for all PubMed publications and all BIMS projects, respectively. For the curious, the scores assigned to individual words matched between BIMS projects titles and PubMed article titles are provided in the spreadsheet *word.score.data*. For each BIMS-PubMed pair, the word score was the sum of the scores of words appearing in both titles.

We need you to do the following:

1. Fill in the *valid.match* column of *pair.data* with 1 for a correct match and 0 for an incorrect match. Leave those pairs of which you are uncertain blank. Here are some suggestions on how to efficiently perform this task.
   1. Set a filter on the *dept.author* column (column I) to contain your name.
   2. Sort by *best.score* in decreasing order and then by *match.score* in descending order.
   3. This will show data only for your publications. Also, the data will be sorted by publication and then by matching BIMS project according to word match score.
   4. Scroll down the list and fill in 1 for correct matches and 0 for incorrect matches.
2. Fill in the BIMS number for unmatched publications in the spreadsheet *unmatched.pubs*.
   1. Set a filter on the *authors* column (column D) to contain your name.
   2. For each paper, search for the BIMS project in the spreadsheet *all.BIMS.data* by using filters for certain terms in the title or certain people in the sponsor, requestor, or initiator columns.
   3. After finding the BIMS number, fill it in the *BIMS* column of *unmatched.pubs*. If more than one BIMS project matches the publication, separate unique BIMS identifiers with a semicolon.
3. Fill in the publication PMID for unmatched projects in the spreadsheet *unmatched.BIMS*.
   1. Set a filter on the initiator column (column F) to contain your name.
   2. For projects with no publication fill in 0. For other projects with a publication, search the spreadsheet all.PubMed.data for the paper. Use filtering and word searches. Fill in the PMID of the matching publication. If more than one publication matches a project, separate unique PMIDs with semicolons.
4. Be mindful that everyone is working with this spreadsheet. Please clear filters and save the file when you are done. Also, be aware that someone else may have inadvertently left some filters in place before you opened the file. You may want to clear filters before beginning your work.