Good afternoon – I’m Ed Sperr, Clinical Librarian from the Augusta University/University of Georgia Medical Partnership in Athens, Georgia.

If you’ve spent some time searching PubMed, you know that what you get back when you perform a search is a long, paginated list of results. This works well for sifting through and selecting individual citations, but it doesn’t do is give us much of a sense of what our search is actually doing.

This project, Visualizing PubMed, is an attempt to display different aspects of a search and give some sense of what is going on beyond just the number of results. All the applications I’m going to demo, as well as the source code underneath, can be found at the URL at the bottom of this slide, but before I get started, I want to talk a little bit about how it’s done.

The most important design decision was to do as much as possible in the web browser. The combination of a modern browser and JavaScript is a powerful one, particularly when we query external information resources through an Application Programming Interface, or API. For Visualizing PubMed, I am using the NCBI E-utils API, which allows one to search the whole suite of NCBI databases. I do want to note that NCBI now has some nice documentation for E-utils at this location if you want to get started yourself.

Another design decision was to focus on interactive searching. With this tool, we can modify our search and see the differences almost immediately. So, one could use this tool not only for visualizing a new search, but potentially for helping fine tune an existing search step by step. Finally, we’re not going to recreate the wheel. Once we get some data back from E-utils, we’re going to make sure that we use existing visualization libraries as much as possible.

So how this works in practice is that we take a search and break it up into different aspects. We then use E-utils to search each of those aspects and count what we get back. Since what we care about is what makes our particular search unique, we then compare that to the baseline of PubMed as a whole and graph whatever differences that we find.

In this case, using MeSH Category Graph, we’re looking at the search “evidence based medicine” and seeing the proportional distribution of top-level MeSH headings, like “Anatomy”, “Health care” and so on. But once we have established this basic pattern, we can expand it to other ways of breaking down a topic, such as in this case, by the proportion of floating subheadings that show up when we search for breast cancer...

...or by the proportion of results for “stomach cancer” in different countries in Asia...

...or even by how the proportions of citations change over time, as in this search for “cholesterol”.

Once we have more than one search, we can think about not only how do these searches differ from PubMed as a whole, but we can look at how they differ from each other as well. We take those proportions we’ve already computed, and we find some way to display the differences.
So in this example, we see that “chemicals and drugs” is proportionally more important for “breast cancer” than it is for “uterine cancer”

....and that proportionally, citations for “beer” are more common in Germany while “wine” is more important in Spain.

...we can see how citations for “malaria” are becoming proportionally less common, with “dengue” steadily increasing, and “Zika” publications exploding since its re-emergence

...or finally, we can take a really broad overview and see how different sets of results overlap using Venn diagrams. Not surprisingly, “death” is a lot more common in PubMed than “taxes”, and there is just a little overlap between sets.

Thank you for your attention.