

Comment

Meta-morphosis

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.

Email: petsko@brandeis.edu

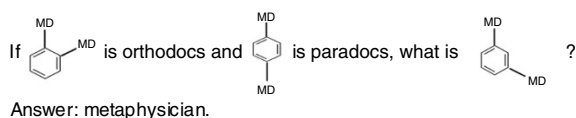
Published: 3 November 2008

Genome Biology 2008, **9**:111 (doi:10.1186/gb-2008-9-10-111)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/10/111>

© 2008 BioMed Central Ltd

There's an old chemistry joke, based on the *ortho*, *meta* and *para* nomenclature for substituents attached to a benzene ring, that goes like this:



OK, I know it's not very funny, but how many funny chemistry jokes are there? Anyway, someone reminded me of this joke the other day and while I was busy not laughing I was also thinking about something else. Because these days, for most people in the life sciences, and I bet nearly everybody in genome biology, the prefix *meta* doesn't conjure up images of di-substituted benzene. It calls to mind the meta-analysis of data.

Meta-analysis - sometimes contracted to metanalysis - is one of those Sudoku-like fads that seem to pop up overnight and sweep the entire country in about as little time. 'Metanalysis' doesn't mean the same thing to scientists as it does to other academics. In linguistics, metanalysis is the act of breaking down a word or phrase into segments or meanings not original to it. The term was coined by the linguist Otto Jespersen, and comes from the Greek for 'a change of breakdown'. Here's an example, courtesy of Wikipedia: in the phrase "God rest ye merry gentlemen", originally merry was a complement with rest (as in "God rest ye merry, gentlemen" - note the position of the comma - that is, "[may God] give you gentlemen a pleasant repose"). But now, by a process of metanalysis, merry is frequently construed as an ordinary adjective modifying gentlemen ("God rest ye, merry gentlemen") - and in all probability it can be relexicalized with the current sense of merry, that is, cheerful, jolly, though that is harder to be certain of. (The expression "to rest merry" and the like was once generally current, by the way.) Incidentally, I love this sort of thing, but it isn't what 'meta-analysis' means in biomedical research.

Meta-analysis for us is a technique whereby all data from all available studies of something are combined, often regardless of the relative quality of the data. The method is used by researchers to get a maximum amount of statistical information from a set of studies that might not have large enough individual sample sizes, or whose results may be of marginal statistical significance.

In a typical meta-analysis, the results of, say, four different clinical trials of a drug, or the data from five independent studies of the association of a genetic variation with a particular disease, are merged into a single statistical sample. The sample is then analyzed for the same correlation, or lack thereof.

I don't know where the practice came from but it's of relatively recent origin - the late 1970s or so. Certainly when I was a student there wasn't an entire subfield devoted to pooling studies and reinterpreting the results. But the medical literature, and the literature of human genetics, is awash with it now.

Imagine the excitement of the person who thought this up. "Omigod! I can take other people's results, throw them together, and come up with a conclusion that in some cases will horrify them, and I can get it published in good journals - sometimes with quite high visibility if the data concern an important human disease or a dietary substance that's popular - and I don't have to do any of the really hard work! I don't have to actually do the studies. I don't have to find the patients, or collect the questionnaires, or analyze the genomes, or any of that difficult, real science. They'll do it for me, and then I can use their data for my own benefit. This is like making money with other people's money! This is how the first banker must have felt!!"

I am not a professional statistician, but I have taught statistics, and as a protein crystallographer I have to be familiar with most of the rudimentary forms of statistical analysis of data. And everything I know about the subject suggests to me that

meta-analysis could lead to all kinds of trouble. One of the things I've learned is that you don't improve bad data by combining it with good data. In fact, exactly the opposite occurs: the bad data degrade the better. "When in doubt, leave it out" is what I usually tell my students - occasionally modified to "when in doubt, weight it down". But most meta-analyses don't apply relative weights to the studies they combine (it's pretty hard to see where they would get the weights anyway), so studies of all sorts of varying quality are sometimes pooled as though they were equally reliable.

And even if we allow that the people who do this may get better statistical precision this way, owing to an enlargement of the sample size, I don't think statistical precision is the issue here. Adding more data may reduce the random errors, but I believe that in most biomedical and genomics studies - especially the latter - the real importance may lie in the systematic errors, or, more precisely (pun intended), the systematic differences.

Meta-analysts justify combining data from different studies in part because doing the same experiment in different ways has long been a way to avoid problems caused by non-random or systematic errors. And it is true that using multiple studies that involve different questions or different experimental techniques, with different patient populations and other variations, may allow global trends to emerge from underneath spikes of systematic differences. But there is a danger there. The first meta-analysis I ever encountered, some years ago, concerned a gene I was interested in. It had been reported that a particular polymorphism in this gene was associated with a reduction in risk for certain diseases in the Han Chinese population. Several other studies, including some with much smaller sample sizes, had looked for a disease association in other ethnic groups but had failed to find any. The meta-analysis pooled all of these studies and concluded, rather definitively, that the particular variant did not confer a change in risk for any of the diseases in question. And their overall statistics certainly bore that out.

There was just one problem: I had read the original Han Chinese study rather carefully, and the work was well done, and the statistical correlation with disease risk was absolutely significant. But if you only read the meta-analysis - and I've seen that referred to repeatedly since and even quoted from at meetings - you wouldn't know that.

Because in the age of genomics, it's not about the general population anymore. Personalized medicine is coming, and our studies of haplotypes and other natural variations make it clear that, even if we can't quite get down to the level of the individual yet in all studies, the genetic (and environmental) background of the population being evaluated can make a huge difference. Ethnic and geographical differences aren't not systematic errors to be averaged out; they're essential components of the ways genes (or drugs, or nutrients)

interact with the human body. The right question to be asking in the case of the study I referred to is not whether the polymorphism alters disease risk in the general population, but rather why it does so in the Han Chinese population. What are the particular combinations of other genetic and environmental factors that cause this variant to become associated with a set of diseases? That's where the really interesting science, and medicine, lies. Average different studies together willy-nilly and you run the risk that sometimes you may average out precisely the those variations that provide the clues we are looking for as to how human health really works.

Now you may think that this column is all about trashing meta-analysis, and, to be honest, it started out that way. But then I had a discussion with David Altshuler, a human geneticist at Harvard Medical School and the Broad Institute at MIT, that made me rethink what was about to become a blanket condemnation. He pointed out to me another goal of meta-analysis in current human genetic studies. The key issues, as he put it, are first, the different statistical thresholds needed in discovery science with a low prior (as compared with hypothesis testing in a well-established field), and second, in human genetics the ability to use the rest of the genome as independent tests to assess the matching of cases and controls in human genetics.

Here's how he explained the first issue: in genetic mapping with the goal of initially determining which genes might actually play a role in human disease, there is a very low *a priori* probability of any variant being associated with any risk (on the order of one in a million), and - that is, with an initial goal of determining which genes might actually play a role in human disease risk - one can't say that a *p* value of 0.01 or even 0.0001 is 'absolutely significant'. He's right: most findings of that magnitude turn out to be entirely irreproducible, and are likely to be false positives. But finding consistent results by meta-analysis increases evidence that the null hypothesis is wrong - and confidence that there is a relationship between the variant and risk. Without that confidence, he asserts, the field was awash with wishful thinking about lots of candidate genes that reductionist biologists had found in cells and in animal models, and wanted to believe were 'genetically validated in people' - but that turned out to be noise.

If you look into it, you will find that there is a long history in genetic mapping of setting the right threshold based on the number of tests. In linkage mapping, the LOD score that indicated linkage between a gene variant and a trait was 3.0 - not $p < 0.05$ (LOD stands for logarithm to the base 10 of odds; a LOD score of 3.0 means the likelihood of observing the given pedigree if two loci are not linked is less than 1 in 1000). By contrast, in genome-wide human genetic studies, a *p* of less than or equal to 0.0000001 is typically required for proof of association of a gene with a trait. And while

biologists often like to say that for 'candidate genes' like the one I mentioned above it is more like $p < 0.05$, that threshold has a history of not always supporting reproducible discoveries. It is in gathering enough data to really get the confidence level to where it needs to be that meta-analysis has proven proved valuable in many cases.

The second issue is the quality of the data that goes into meta-analysis, and the ability to compare and align it. In most studies, Altshuler agrees with me that you can't know if you are washing out good data with bad. But he makes a strong case, with which I am forced to agree, that in the current wave of genome-wide association studies, the studies often use the same phenotype definition, or the same microarray protocols, and so the data may well be comparable (clearly, it is important to see whether that is the case, if you want to evaluate the meta-analysis critically). When done properly, information from the rest of the genome can be used to assess the properties of the data, matching of cases and controls, and so on, which, as he puts it, "can result in valid combination of lots of good data, rather than sloppy mashing up of lots of bad data". Finally, sometimes you actually may want to test the hypothesis across ethnicity or age or other variables. In those instances, meta-analysis allows you to put the data together in a valid way - a way that is probably better than just reading the papers and trying to compare them.

Dr Altshuler went on to make a very valuable point, which I think belongs in this column: he feels there that is currently a big and unfortunate divide between cell and molecular biologists and people using human genetics to find disease genes. The tone of the first part of my column just reinforced his concern, I'm sorry to say. For disease research, it's certainly true that knowing up front that a gene influences the disease in humans is invaluable, and I think he's right that meta-analysis can be a powerful tool for getting that correct. (As he asked me: "Would you want to do functional work on a gene that turns out to be a false positive?")

A final point: meta-analysis is currently being used in two ways in human genetics, and I may have given a false impression that there is only one. One method is when there is a pre-existing hypothesis that needs to be tested. A good example comes from Altshuler's own work (Altshuler D, *et al.*: *Nat Genet* 2000, **26**:76-80). The data the authors had collected gave a positive result, but others had claimed the opposite. The studies were comparable enough that meta-analysis was able to show that the data were actually consistent and reinforced one another. In such cases meta-analysis serves as a reality check, and helps avoid possible bias in the selection of data that might support one hypothesis or another.

But careful meta-analysis can also in some cases be hypothesis generating, because enough studies might, as I

mentioned above, allow previously undetected signals to rise above the noise. These can then be tested experimentally, and of course, need to be.

Dr Altshuler closed our discussion with a nice comment: "Meta-analysis," he pointed out, "is just a method. Garbage in, garbage out. But in genetic mapping of common complex diseases, where it is clear there are many different variants contributing, and where studies are expensive, combining data to learn the most is well justified and valid."

It is true that it's often the original studies that matter, that data need to be examined carefully, and that a study ideally must be accepted or rejected on its own merits. People are fond of saying that the devil is in the details. But God is in the details too. In the age of genomics, the details are ultimately what matter: the complex interplay of individual gene and genetic background, diet, environment, perhaps even state of mind, is what determines whether we are prone to this disease or that, react poorly or well to this drug or that, age well or badly. Meta-analysis can hide those details, but used properly, it can also be revealing. I guess you could say that writing this column has led to a meta-morphosis in the way I think about this subject, and that I have to back away from my meta-phor about making money with other people's money. In the right hands, meta-analysis can be a valuable tool, which is a pity, because it's so much more fun to write a column that completely trashes something. Meta-physically speaking, of course.