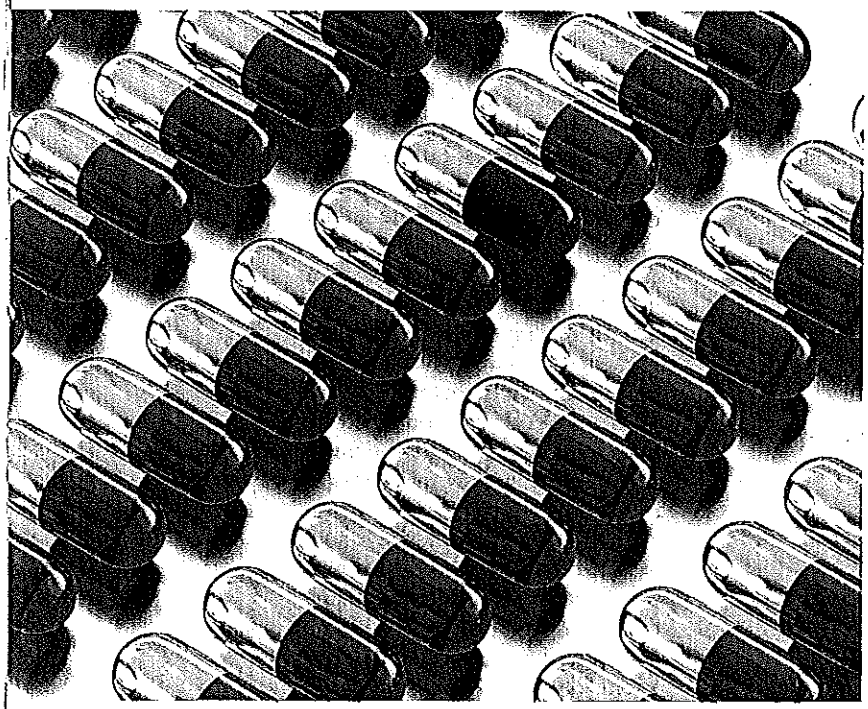
[Ethics and Statistics]

Andrew Gelman
Column Editor

Ethics in Medical Trials: Where Does Statistics Fit In?



For this issue's column on ethics and statistics, I consider two examples of medical trials. The first is a clear case of an ethical violation described by bioethicist Carl Elliott. As a doctor, Elliott focuses on individual patents; as a statistician who has done a small amount of consulting for pharmaceutical companies, I've been trained to focus on the goal of accurately estimating treatment effects. (Medical ethicists such as Ezekiel Emanuel have written about the sometimes conflicting goals of caring for the patients in a study and providing expected benefit to future patients.)

In an article in the *London Review of Books*, Elliott wrote of a medical research outfit that was not merely unethical, but also a criminal enterprise:

> In Miami, investigative reporters for *Bloomberg Markets* magazine discovered that a contract research organisation called SFBC International was testing drugs on undocumented immigrants in a rundown motel; since that report, the motel has been demolished for fire and safety violations. ... SFBC had recently been named one of the best small businesses in America by *Forbes* magazine [in 2003, ranked #3 in the nation]. The Holiday Inn testing facility was the largest in North America, and had been operating for nearly ten years before inspectors noticed there was anything wrong.

Elliott writes of the bigger picture:

> Over the past 20 years or so, without much fanfare, clinical research has undergone a remarkable free-market conversion. ... What does clinical research look like when everyone is in it for the money? For a start, it looks a lot less like science. 'I do not do original research; I do contract research,' says a private physician-researcher in *Medical Research for Hire*. A contract researcher does not come up with original ideas, or design research protocols, or analyse research results, or write them up for scientific publications. All of this is done by the pharmaceutical company or its hired specialists. What a contract researcher does is recruit subjects, monitor their clinical care, and sign off on the paperwork. Not a lot

of original work is done, and in some cases, not much work at all.

I agree with Elliott that the conflicts of interests here are huge. There is no particular reason why the doctors involved in a study should feel they are doing original research; maybe it's better if they are completely focused on monitoring and taking care of the patients. But are these doctors actually taking care of anybody? Or are they just being paid for their MD credentials? It doesn't look good:

> Usually they will 'come in on a daily basis, on most days, and they'll sign off on all the things they need to sign off on, see any patients they need to see, and they're gone. ...' The researchers are usually onsite for no more than an hour or two a day. Contract researchers may not do much intellectual work, but this doesn't mean they are not well paid. A part-time contract researcher conducting four or five clinical trials a year can earn an average of $300,000 in extra income. ... Even an ordinary office visit will be paid at twice the usual rate if the visit is part of a research study. ... Contract researchers may find that their sponsors do not welcome bad news about the trials, especially if the drug appears unsafe. Reporting that subjects have experienced a 'serious adverse event' (industry-speak for the worst side effects) may mean losing the contract.

Sounds like bribery to me. The story, as I take it, is that these doctors are being paid big bucks to keep their mouths shut, to keep people in the study no matter what, and to downplay adverse events. An independent doctor, not dependent on the drug company for the money, might very well advise a patient to take a different treatment if problems arise.

Here, I wonder if statisticians are part of the problem. We go on and on about the threats to validity of causal inference if patients drop out of a study or don't take the assigned treatment, and this puts enormous pressure on researchers to not "cheat" and to keep everybody in. Perhaps recent statistical research on causal inference from broken experiments is not merely helpful, but necessary for ethical experimentation.

There are also familiar statistical problems involving effect sizes and multiple comparisons. Elliott writes that many studies are of drugs that are only "incrementally better than a control drug or placebo." If the benefit is small, sample sizes need to be larger, which is not good for the benefit-cost ratio for the trial. He also writes that many studies are done in other countries "when their trials in the West have failed to show the drug is effective." No amount of statistics on an individual trial will substitute for a data repository and meta-analysis.

Avastin: What's the Story?

My next example is more complicated. Avastin is a cancer drug produced by Genentech that Medicare pays for in breast cancer treatment. According to *New York Times* columnist Joe Nocera, "Avastin neither suppresses tumor growth to any significant degree nor extends life" for breast cancer patients. As a result, the FDA decided to not approve the drug. Despite this ruling, Nocera reports, Medicare continues to pay for Avastin for breast cancer patients at a cost of nearly $90,000 a year per patient.

After seeing this in the newspaper, I wondered if there was more to the story, so I asked some of my colleagues who work in medical statistics. Psychologist Jon Baron recommended a formal decision analysis using quality-adjusted life years. Such calculations can be controversial, but they have the virtue of putting the assumptions and tradeoffs right out in the open, where they can be discussed and debated. Epidemiologist Sander Greenland remarked that the key issues involved politics and economics, rather than statistics, in that there is a conflict of interest when drug companies and advocacy groups are involved in Medicare decisions. And, giving a comparative perspective, Australian biostatistician John Carlin reported, "Avastin (bevacizumab) is still approved here as a treatment for certain indications, but it is not approved for listing on the Pharmaceutical Benefits Scheme, and without receiving a subsidy via that mechanism, its cost is generally prohibitive."

A prominent U.S. health care economist pointed out an interaction between statistical and political concerns, writing:

> Avastin is believed to work for some women, but not all. No one knows which women will benefit, and so the FDA ordered Roche/Genentech to figure this out. The technical issue is that, by law, CMS has to cover any approved cancer drug for an off-label indication if the indication is included in one of several named drug compendia. ... The National Comprehensive Cancer Network (NCCN), 'a non-profit group of oncologists whose guidance is closely followed by leading treatment centers, has voted overwhelmingly in favor of maintaining its recommendation that Avastin should be used to treat breast cancer.' The NCCN vote was 24–0, with one abstention. That brings up the politics here ... how organizations like the NCCN make their decisions ... a bunch of members of the NCCN have ties to Roche; they make their money off this stuff; etc.

Finally, statistician Don Berry, who is an expert on medical decisionmaking in general and cancer treatment in particular, writes:

> There is no question that Avastin 'works' in the sense that it has an anti-tumor effect.... It clearly delays progression of metastatic breast cancer, which is the reason it was approved for treating that disease in the first place. The FDA reversed itself (for this disease, but not for other cancers such as lung and colon that will remain on Avastin's label) because Avastin has not been shown to statistically significantly prolong overall survival (OS). Some oncologists—actually, most oncologists—argue that progression-free survival (PFS) is clinically meaningful and should be a registration end point. The FDA's position—and that of the Oncology Drug Advisory Committee—is that improved PFS is not usually enough to approve drugs without empirical evidence of improved OS to go along with it.

Berry then connects to some familiar ideas from statistical design and analysis:

> It's very difficult to power a study to show an OS benefit when survival post-progression (SPP=OS-PFS) is long, which it is in metastatic breast cancer—about two years in some of the Avastin trials. ... Even if an advantage in PFS translates perfectly into the same advantage in OS, the variability after progression so dilutes the OS effect that it's likely to be lost.... I know many examples of clinical trials in many types of cancer (and I know no counterexamples) where SPP is essentially the same in both treatment groups, even when the experimental drug showed better PFS than control. This is despite crossovers (to the drug when the control patient progresses) and potentially greater efforts by the clinicians in one treatment group to keep their patients alive. (The main reason that SPP is similar in the two treatment groups is that metastatic cancer is almost uniformly fatal and it's hard to slow the disease after it's set up housekeeping throughout the body.) It makes sense that a drug that was effective in delaying progression is not effective after progression, because the drug is almost always stopped when the patient progresses, and the patient usually goes onto another drug.

What are the lessons from these examples? How can we as statisticians contribute (or, at least, not make things worse)? Given that we as a field can't even agree on methods, we certainly can't be expected to come to a common conclusion on any particular question of design or data analysis—and that's not even considering the distortion of financial and professional incentives. We can, however, move toward openness (most simply, the archiving of all results, not just those that are statistically significant) and toward mainstreaming of methods for analysis of difficult data. (Recall the doctors who are pressured to keep patients in the study and to suppress reports of adverse outcomes.)

It also can make sense to move toward a cost-benefit framework for decision analysis, rather than to let the analysis stop at reports of confidence intervals and statistical significance. It is not necessarily appropriate for a statistician to perform the decision analysis, but we should provide inferences in a form for which they can be incorporated into such a calculation.

Finally, I think we need to question the assumptions that underlie many conventional analyses. Consider, for example, Berry's discussion of progression-free survival instead of overall survival as a cancer-trial endpoint. Biostatistician John Johnson writes:

> Overall, survival is seductively simple because it's a 'hard' endpoint (few people disagree on when a person dies), as opposed to PFS or anti-tumor activity, both of which involve some degree of subjectivity. (PFS can also be a pain to compute if you don't do your data collection very carefully.) Add to this the issue that anti-tumor activity doesn't always translate into long-term benefit because tumors can build up resistance much the same way that bacteria can build up antibiotic resistance. (And so patients switch drugs upon progressing and complicate the OS endpoint.)

Ultimately, there is no simple safe statistical analysis, any more than there is a safe medical treatment. It's best if we're open and as quantitative as possible about the tradeoffs. And "quantitative" is something statisticians should be good at.

Further Reading

Joe Nocera. 2011. Why doesn't no mean no? *The New York Times* Nov. 21, Page A29.

Broglio, K.R., and D. A. Berry. 2009. Detecting an overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute* 101:1642–1649.

Elliott, C. 2010. The mild torture economy. *London Review of Books* 32(18):26–27.

Emanuel, E. J., D. Wendler, and C. Grady. 2000. What makes clinical research ethical?" *JAMA* 283(20):2701–2711.

About the Author

Andrew Gelman is a professor of statistics and political science and director of the Applied Statistics Center at Columbia University. He has received many awards, including the Outstanding Statistical Application Award from the American Statistical Association and the award for best article published in the *American Political Science Review*. He has coauthored many books; his most recent is *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*.