Which Assay to Believe?

My PI and I were working on an experiment to see if Y occurred when a particular gene was knocked out. If Y did indeed occur, we would be keen to publish the finding. Determining whether or not Y occurred would require our doing some assays.

The problem was that assay #1 confirmed Y, but assay #2 disconfirmed Y. It must be noted that these were not duplicate assays but completely different ones. And repeated assays of both types kept giving us the same contradictory findings.

So, the first issue was the temptation to simply tell my PI of the preferred result, namely the one that confirmed Y. I told myself this wouldn't be a lie, but it came close. So, I told the PI about both assays. At this point we were confronted with the following decisions: Should we publish (and therefore believe) the results from the assay that worked and disregard the "bad" one, assuming that the problem involved some flaw in that assay? Or should we mention in the publication that we were only able to show Y with one assay, and not with another? Or should we try a third assay and go with it (as a tie-breaker)?

We went with the last strategy which fortunately confirmed Y and justified our paper. But suppose there wasn't a third assay available? What then?

Expert Opinion

The literature calls the kind of temptation the dilemma contributor describes an instance of "selective reporting."[1] Although not exactly a manipulation of data, selective reporting is more a kind of "editing" one's findings such that the data that are reported put those findings in the best possible light. While the above case concerns the temptation to withhold disconfirming test results, other instances of selective reporting might involve deleting data points, succumbing to pressures by a commercial research sponsor to report on only one aspect of a study (e.g., superior outcomes rather than worrisome adverse events), using percentages rather than actual numbers (so as to omit mentioning that the actual sample size was very small), applying multiple statistical tests to the same data set but only reporting the test(s) that yield the most favorable results, and ignoring mention of prior research that challenge the stated findings.[2]

One cannot help but recall the Vioxx scandal, which began with the publication of the VIGOR study in the *New England Journal of Medicine* in November 2000.[3] The authors of that study omitted their finding that Vioxx carried a five-times higher risk for myocardial infarction, thrombo-embolic events, hypertension and heart failure than the comparator, naproxen.[4] By the time the VIGOR study appeared, however, the FDA had already approved Vioxx despite knowing about the elevated cardiovascular risks. (Apparently, the FDA accepted the investigators' rather remarkable explanation that Vioxx's apparent risks were actually due to "a potential cardioprotective effect of the comparator drug."[4, p. 348]) By 2002, however, the FDA had received enough reports of adverse cardiovascular events associated with Vioxx to prompt requiring Vioxx's manufacturer, Merck Sharpe & Dohme, to list those risks in Vioxx's package insert. By 2004, when the APPROVe (Adenomatous Polyp Prevention On Vioxx) study appeared and confirmed Vioxx's dangerous risk profile, Merck voluntarily withdrew the drug worldwide.[4]

Our reason for recalling the Vioxx case is that its lessons speak directly to the above case: The reporting of research or clinical results must contemplate other investigators or clinicians repeating those experiments or interventions. Investigators who simply want to publish findings that support their hypotheses suffer from a kind of professional or ethical myopia. Did the Vioxx investigators really think that the drug's side effect profile would go

unnoticed ad infinitum? If they believed their research was competently performed, did they think that Vioxx's consumers wouldn't begin demonstrating precisely the kinds of cardiovascular symptoms that the study participants did? Vioxx is perhaps an extreme case, but that's why it's worth remembering. The consequences of selective reporting can be catastrophic: Consumers profoundly harmed, professional careers trashed, the public's trust in science seriously eroded, and extremely costly litigation waiting in the wings.

Specific to the case above, bench scientists would likely assert that one ought never rely on only one experimental approach.[5] Indeed, very rarely is any one assay result definitive. Because it can be maddeningly difficult to control for all the variables that can affect a result or a finding, one generally wants to test one's hypothesis in as many ways as one can. The research team that selectively reports data from a single assay will likely arouse the suspicions of any competent reviewer, who will wonder why other assays weren't performed.

This will especially be the case if the investigator's experimental question has a significant history. That history will probably frame or suggest the number and kinds of assays the scientific community will expect to be reported, enumerate the variables to be controlled, describe the potential for misinterpreting findings (e.g., sometimes an experiment succeeds but not necessarily for the reasons the investigators posit) and suggest which data to believe, which to doubt, and which to report.

Nevertheless, there can be considerable value in reporting results that are inconsistent as well as consistent with the hypothesis. In the above scenario, if a third assay wasn't possible, the investigators would have done well by their colleagues to have reported the results of <u>both</u> assays. That way their peers will have a truthful and complete rendering of the experiment and its results, whereas a partial or selective presentation of data slows the engine of science: Other researchers will have to discover the partiality of the data, call professional attention to it, and fill in the gaps—all of which can take a great deal of time. (This is one reason why reporting negative data can be so valuable, and why investigators often lament the apparent journalistic bias against it.)

Investigators who encounter disconfirming as well as confirming data might greatly benefit from peer advice and recommendations on how to present such findings. One question to ask, for example, is whether or not there is historical precedent for the discrepant assay results and what the explanation might be. We are not told in the scenario what kinds of assays are being carried out, e.g., in vivo or in vitro; nor can we identify certain experimental conditions that might account for discrepancies, e.g., temperature, the use of a particular dye or stain, etc.; nor are we told about the finding of interest, e.g., a cellular structure or a behavior. Of course, it might be the case that the reason for the negative or disconfirming result is a faulty design or errors in the assay. Unlike clinical lab determinations that are done thousands of times and, one hopes, have reasonably good of quality control, this is often not feasible for most research lab determinations.

In any event, because assays are unnatural intrusions into natural processes, the more that investigators deploy multiple approaches that control for those variables (and possible errors) and that can support results, the more those results will appear confirmed. To reiterate: If an obvious, confirmatory test is not carried out, the investigators should be prepared to explain why.

Ultimately, selective data reporting retards the efficiency and momentum of scientific discovery; it can waste huge amounts of money; and, as the Vioxx case illustrated, it can pose great harm to research participants and health care consumers. As Marco and Larkin have pointed out, scientific research that is publicly funded but that is incompetently or unethically performed is a violation of the social contract in research.[2] That contract at least anticipates the

truthful and honest reporting of research data in return for the funds that make the experiments possible. The investigator who knowingly and intentionally reports only that fraction of his data that puts his experiment in the best light has reneged on the social contract and has fallen victim to his self-interests. And because scientific research and clinical care are inevitably self-policing, succumbing to such selfish inclinations might not only harm the public that scientific research is supposed to benefit, but end the careers of otherwise talented and hard-working investigators.

References:

1. Fielder JH. The Vioxx debacle revisited. *IEEE Engineering in Medicine and Biology Magazine*. July/August 2008;106-109.

2. Marco CA, Larkin GL. Research ethics: Ethical issues of data reporting and the quest for authenticity. *Academic Emergency Medicine*, 2000;7:691-694.

3. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ. VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *New England Journal of Medicine*, 2000; 343:1520-1528.

4. Cahana A, Mauron A. The story of Vioxx—no pain and a lot of gain: Ethical concerns regarding conduct of the pharmaceutical industry. *Journal of Anesthesia*, 2006;20:348-351.

5. Personal communication with Professor Arri Eisen, Emory University, Feb.13, 2009.

©2009 Emory University