VIEWPOINT

Rethinking Reanalysis

Dimitri A. Christakis, MD, MPH
Seattle Children's Research Institute, Seattle, Washington.

Frederick J. Zimmerman, PhD
UCLA School of Public Health, Los Angeles, California.

Data sharing is becoming increasingly common, whether at the insistence of the National Institutes of Health (NIH) compelled by freedom-of-information requests, or because of congressional subpoena. The increasing ease with which data can be acquired by those who did not originally collect them has led to an emerging scientific form, the reanalysis, in which a new team uses previously collected data to answer a question previously addressed by the same data. Reanalysis is different from replication, in which new data are used to answer an old question, and from supplemental research, in which previously collected data are used to answer a new but related question.

The rationale given by the NIH to support data sharing in 2003 and reaffirmed in 2013 is that it is "essential for expedited translation of research results into knowledge, products and procedures to improve human health."[1] This is true in the case of supplemental research; however, data sharing for the purposes of reanalysis, although it may have prima facie appeal, does not necessarily serve the public good and may be counterproductive unless conducted within certain constraints.

The potential value of reanalysis was highlighted in a highly publicized case of recombinant bone morphogenetic protein-2 (rhBMP-2).[2] In that case, the safety of the drug was questioned after several industry-sponsored studies were alleged to have overstated benefits and minimized risks.[2] In response to the public uproar, Medtronic agreed to release its data to researchers at Yale University for an independent analysis in what has been appropriately billed as a model program of unbiased analysis of industry data.[3] The resulting meta-analyses, pooling all available data, concluded that rh-BMP-2 was no better than traditional treatment and that it was associated with increased postoperative pain.[4,5] This episode appears to make a convincing case for reanalysis and for the data sharing on which it depends. Yet this was not just any reanalysis. In this case, there were clear reasons for concerns about the original science because the funder was an international pharmaceutical company that stood to profit on the basis of the studies' outcomes, and study authors were alleged to have received millions in royalties and fees.[6] On the other side, the parties conducting the reanalysis were experienced and unbiased investigators acting as members of the Yale University Open Data Access Project in the interest of public health. Yet instead of implying a rush to reanalysis, this case is a cautionary tale about the potential pitfalls of data sharing.

> The effect of burgeoning reanalyses is to create a scientific landscape in which multiple studies using the same data yield disparate results.

The heart of reanalysis is to use data from a study that has found a certain outcome and scrutinize those data, in effect using the results of the original analysis as the new null hypothesis. A reanalysis uses the same data to confirm or refute prior findings. For example, as with rhBMP-2, the outcome of interest could be the effectiveness of a particular drug or its adverse events. However, it could also be the adverse effects of a chemical agent, the association between insurance coverage and health, the benefits of mammography screening, or the relationship between television viewing and child development.

A major threat to the validity of reanalysis comes from the potential for increased investigator bias. In the case of rhBMP-2, the financial interests of the funder of the original study led to a plausible presumption of bias, whereas those seeking the data for reanalysis were plausibly free of bias. In other reanalyses these identities could be reversed. For instance, a recently published study on the relationship between diesel exhaust and cancer outcomes,[7] conducted by carefully chosen independent researchers, was subjected by the diesel industry to repeated requests for the raw data and attempts to thwart the project.[8] Masquerading under the name Methane Awareness Resource Group, a coalition of mine operators had successfully delayed publication for more than 20 years. The point was not to advance science but to slow regulation. Now that the initial report has been published, a reanalysis may soon follow. In a currently active example, House Republicans, possibly motivated not by science but by politics, have demanded the raw data for 2 peer-reviewed studies published in the early 1990s that linked air pollution to adverse health effects.[9] The implications for science of this kind of reanalysis are potentially chilling.

The value of reanalysis accordingly hinges critically on reducing the presumed threats to equipoise that come from a financial, ideological, or political interest in the results. A reanalysis is most likely to be useful when such interests are substantially lower among the reanalysis team than in the original team. Conversely, if the presumption of bias is higher in the reanalysis team, data sharing will more likely impede, not improve, scientific understanding.

Furthermore, because the universe of researchers with the expertise, time, and interest in reanalyzing another researcher's data can be quite small, there is a wor-

Corresponding Author: Dimitri A. Christakis, MD, MPH, Department of Pediatrics, University of Washington, 1100 Olive Way, Ste 500, Seattle, WA 98101 (dimitri.christakis@seattlechildrens.org).

risome likelihood that those seeking to reanalyze data either have a vested interest in refuting the findings or do not have adequate methodological expertise. Because methodology drives the quality of results, this is not an idle concern. Anyone sufficiently motivated can produce a different and conflicting result based on data that once demonstrated a given outcome. A misspecified model, a reduction in power, inappropriate controls; the pitfalls are many and one investigator's pitfall is another's opportunity.

Overlaying the potential for investigator bias is the inherent publication bias in reanalyses. For most studies, a reanalysis that reaffirms prior published results would contribute nothing and hence be considered unpublishable. As such, only reanalyses that arrive at different conclusions from prior work are likely to be published, leading to perverse incentives for those reanalyzing data to refute prior studies even if they were not already predisposed to do so.

The effect of burgeoning reanalyses is to create a scientific landscape in which multiple studies using the same data yield disparate results. To many in the scientific community and to the public at large, the appearance that a reanalysis refuted the original results may be perceived as the final word. The original investigators are then left only with attempting to conduct a rereananalysis to verify their original findings which, even if persuasive to experts, would likely appear to be defensive argumentation.

The result of uncritical reanalyses is not better science but scientific cacophony. Worse still, given the publication-bias problem, a false equivalency could be created with the appearance that there are 2 groups of scientists in disagreement, whereas in fact there might be near universal consensus within a group of experts but 2 published, contradictory papers. The resulting contentious science does not promote public health and can easily be exploited by those with political or financial objectives.

Box. Core Principles of Study Reanalysis

A reanalysis should not be a statistical fishing expedition. The new methodological approach must be explicitly stated and justified in advance. The precise question and methodological approach should be described in a prespecified protocol, even recorded at a clearinghouse site such as clinicaltrials.gov, before the data are released to the new investigator.

The presumption of bias arising because of financial, ideological, or political interests should be at least as low in the reanalysis team as in the original team, and ideally lower.

The methodological improvement should be recognizable and significant. While there are legitimate methodological differences, the posited improvement from the reanalysis should be well grounded and substantiated by a significant portion of the methodological literature.

The authors of the original report being subjected to reanalysis should be provided with the opportunity to review and comment on the reanalysis before its acceptance for publication.

To help ensure that reanalyses serve their intended and worthwhile function, several key core principles should be incorporated into any reanalysis (**Box**).

With such safeguards in place, reanalysis could play its role in enhancing scientific discourse, but embarking on reanalyses without these caveats may jeopardize science and public health.

ARTICLE INFORMATION

Conflict of Interest Disclosures: Both authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Christakis reports receipt of grants from the National Institutes of Health that are unrelated to this article. Dr Zimmerman reports no disclosures.

REFERENCES

1. National Institutes of Health. Final NIH statement on sharing research data. http://grants.nih.gov/grants/policy/data_sharing/. Accessed September 25, 2013.

2. Meier B. Medtronic giving Yale grant to review bone growth data. *New York Times*. August 4, 2011:B5.

3. Krumholz HM, Ross JS. A model for dissemination and independent analysis of industry data. *JAMA*. 2011;306(14):1593-1594.

4. Simmonds MC, Brown JVE, Heirs MK, et al. Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. *Ann Intern Med*. 2013;158(12):877-889.

5. Fu R, Selph S, McDonagh M, et al. Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis. *Ann Intern Med*. 2013;158(12):890-902.

6. Weaver C. Studies fail to back Medtronic spine product. *Wall Street Journal*. June 17, 2013.

7. Attfield MD, Schleiff PL, Lubin JH, et al. The Diesel Exhaust in Miners study: a cohort mortality study with emphasis on lung cancer. *J Natl Cancer Inst*. 2012;104(11):869-883.

8. Rosenstock L, Lee LJ. Attacks on science: the risks to evidence-based policy. *Am J Public Health*. 2002;92(1):14-18.

9. Rowland C. House GOP demands Harvard study data. *Boston Globe*. September 7, 2013.