

Epidemiology Commentary

What can we really learn from observational studies? The need for empirical assessment of methodology for active drug safety surveillance and comparative effectiveness research

David Madigan, Columbia University, OMOP, and Mini-Sentinel

Patrick Ryan, Johnson & Johnson and OMOP

Corresponding Author:

David Madigan,

[Department of Statistics,](#)

[Columbia University,](#)

1255 Amsterdam Ave.,

New York, NY 10027, USA

madigan@stat.columbia.edu

(212) 851-2131

Running Head: What can we really learn from observational studies?

Acknowledgements: Bram Hartzema, Judy Racoosin, Christian Reich, Paul Stang, Emily Welebob,

In 2007, the United States Congress passed the Food and Drug Administration (FDA) Amendment Act, which called for the establishment of an “active postmarket risk identification and analysis system” with access to data from 100 million lives by 2012.¹ The American Recovery and Reinvestment Act of 2009 committed \$1.1 billion to comparative effectiveness research, to enable “the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in “real world” settings”.² Both developments rely on expanded secondary use of observational healthcare data and have thrust the field of epidemiology into the national spotlight. In particular, these efforts require a significant evolution in the use of these data from the customized design of an individual study of a particular product-outcome association to a broader effort that effectively uses these data for active monitoring of any medical product and any health outcome of interest across a network of disparate databases. The envisioned systems would go beyond the retrospective evaluation of hypothesized effects to proactively explore the data to generate and refine hypotheses of potential issues and benefits that warrant further scrutiny. Little research currently exists to inform the development of such systems.

Focusing just on active drug safety surveillance, open questions include:

- Which statistical and epidemiological methods work best for different types of drugs and outcomes?
- What is the optimal way to combine information from disparate databases?
- What thresholds should active surveillance systems use to identify risks?
- What processes should exist to handle newly identified risks?
- How often will active surveillance system yield false positives and false negatives?

The last question is key and no answer to it currently exists. In fact, the absence of established operating characteristics bedevils all of pharmacoepidemiology. Typical studies in observational databases attempt to control for bias by careful control for confounders, restrictions to certain patient subgroups, etc. While these issues have received scrutiny through academic discourse and isolated anecdotes in the literature, no systematic studies exist that elucidate how these issues manifest themselves in real-world data. Consequently, no assurance can be provided that estimated relative risks, odds ratios, etc., that these studies produce are anywhere near the corresponding true values or that standard 95% intervals actually contain the truth 95% of the time.

The practical complexities and intrinsic biases in real-world data suggest that answers to these questions are unlikely to arise from theoretical research and instead must derive from extensive empirical experimentation. The Observational Medical Outcomes Partnership (OMOP), a public-private partnership chaired by the FDA and managed through the Foundation for the National Institutes of Health, represents one step in this direction. OMOP conducts methodological research to inform the national drug safety efforts by empirically measuring the performance of an array of alternative analysis methods across a network of 10 databases covering over 200 million patient lives³.

OMOP studies the operating characteristics of methods that attempt to discern true drug-outcome associations from background noise. The basic approach thus far has comprised three bodies of work:

1. Establishing a data network, comprising the types of data anticipated for use in active surveillance and comparative effectiveness research, both administrative claims and electronic health records covering different populations, and converted to a common data model format.

2. Creating a research laboratory to enable methodologists to develop and implement analytical approaches for estimating the effects of medical products. To date, collaborating organizations that participated in methods development activities have resulted in 14 computationally feasible approaches. This included systematic implementation of traditional epidemiologic designs, such as propensity-score adjusted inception cohort designs, case-control surveillance, case-crossover, and self-controlled case series.
3. Applying methods to specific drug-outcome pairs across all databases within the OMOP data network. In the initial experiments, OMOP used 53 drug-outcome pairs as ground truth with 9 true positive drug-outcome pairs and 44 negative controls. Product labels, prior observational studies, and expert consensus informed this classification. A variety of metrics such as sensitivity, specificity and area under the curve quantified the extent to which estimates cohered with the ground truth.

In accordance with its mission, all work products developed throughout OMOP are publicly available on the OMOP website (<http://omop.fnih.org>), including methods descriptions and source code, data characterization tools, white papers, and public presentations. OMOP released the findings of its initial experiments at its 2011 OMOP Symposium, and presentations and audio narrations are available.

OMOP was initially conceived as a two-year research project, aimed at conducting a series of experiments to measure methods performance. The research has raised additional questions that require further investigation. As the research progressed, several groups recognized the value of sustaining a shared resource to facilitate collaborative methodological research across all stakeholders. As such, OMOP continues to serve this vision and has invited those interested in collaborating to join the broader OMOP community. We encourage interested researchers to

register on the OMOP website (<http://omop.fnih.org>) and contact the OMOP team if they are interested in actively participating in this effort. Ongoing work includes exploring the existing results to understand why current approaches yield false positives and false negatives, and developing strategies to overcome these limitations.

The OMOP experiments represent an initial attempt to empirically estimate the operating characteristics of standard pharmacoepidemiology procedures applied to observational data. However, OMOP is not alone in the conduct of methodological research to inform the appropriate use of observational databases. European efforts to assess the performance of active surveillance methods across international data sources include IMI-PROTECT⁴ and EU-ADR⁵. The Mini-Sentinel initiative in the US has initiated multiple workstreams examining different classes of methodology (see <http://mini-sentinel.org>). In the context of vaccines, the vaccine safety datalink project has presented initial findings⁶ showing that their current approach yielded one true positive from ten putative signals, a positive predictive value of 10%.

While early results from OMOP and from these various efforts demonstrate the promise of systematic observational analyses, they highlight significant methodological challenges that are faced when attempting to produce unbiased effect estimates in observational healthcare databases. The results also suggest much work is needed to improve the performance of existing epidemiologic methods, to maximize the likelihood of observing true effects while reducing the risk of false positive findings. Specific areas of opportunity include: enhancing strategies to identify and mitigate sources of bias, improving the extraction of clinical elements from the underlying data, developing solutions for refined outcome definitions, establishing strategies to integrate observational results across disparate data sources, and designing a broader framework for how observational analyses can contribute to a causality assessment.

The empirical findings to date appear to be consistent with prior concerns about the accuracy of epidemiologic investigations⁷, including the suggestion that only large effects be considered as credible from observational analyses⁸ and the occurrence of inconsistencies in different published analyses of the same drug safety concern^{9,10}. Until observational data and analytic methods improve to the point that operating characteristics are consistent with the public's expectations, a credibility gap will continue to exist for epidemiologists working in this area.

Observational studies should serve as a primary, and in some circumstances, best source of evidence in evaluating the comparative safety and effectiveness of medical products¹¹, but only if epidemiologic investigations have sufficient reliability to provide stakeholders with the confidence needed to inform medical decision-making. This involves demonstrating to the skeptics that epidemiology can be more than an art and treated as a science, following systematic processes that generate reproducible and fully transparent research, with well-understood operating characteristics. In particular, clinical insights that currently underlie many design decisions in pharmacoepidemiology need to be applied consistently and subjected to rigorous evaluation.

Advancing the science of active surveillance and comparative effectiveness requires interdisciplinary collaboration between epidemiology, statistics, health services research, computer science, medical informatics, engineering, and the clinical sciences. Substantial research challenges remain but progress can translate directly into significant benefits for patients.

References

1. *Public Law 110-85: Food and Drug Administration Amendments Act of 2007*. 2007.
2. *Federal Coordinating Council for Comparative Effectiveness Research Report to the President and the Congress*. 2009, US Department of Health and Human Services.

3. Stang, P.E., et al., *Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership*. Ann Intern Med, 2010. **153**(9): p. 600-6.
4. IMI-PROTECT. 2011 [cited March 8, 2011]; Available from: <http://www.imi-protect.eu/>.
5. Coloma, P.M., et al., *Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project*. Pharmacoepidemiol Drug Saf, 2011. **20**(1): p. 1-11.
6. http://www.brookings.edu/~media/Files/events/2010/0921_signal_refinement/08_Session%20II_Lieu.pdf
7. Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124.doi:10.1371/journal.pmed.0020124
8. Temple, R. Meta-analysis and Epidemiologic Studies in Drug Development and Postmarketing Surveillance. JAMA, 1999; 281(9):841-844.
9. Green J., et al. Oral bisphosphonates and risk of cancer of oesophagus, stomach, and colorectum: case-control analysis within a UK primary care cohort, BMJ 2010;341:c4444
10. Cardwell, C.R., et al. Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer, JAMA. 2010;304(6):657-663
11. Dreyer, NA, et al. Why observational studies should be among the tools used in comparative effectiveness research. Health Aff. 2010 Oct;29(10):1818-25