Drug Saf (2015) 38:101–103
DOI 10.1007/s40264-014-0252-x

LETTER TO THE EDITOR

Comment on: "Desideratum for Evidence-Based Epidemiology"

Sean Hennessy · Charles E. Leonard

Published online: 16 December 2014

Dear Editor,

We agree with Overhage et al. [1] from the Observational Medical Outcomes Partnership (OMOP) that better reproducibility is needed in pharmacoepidemiology. In addition to epidemiology in general [2], the need for better reproducibility has been highlighted across many scientific fields including psychology [3], economics [4], laboratory-based biologic research [5, 6], genomics [7], computational science [8], clinical trials [9, 10], and meta-analyses of clinical trials [11]. OMOP's goal of rigorously evaluating the approaches and databases used in pharmacoepidemiology is therefore important and laudable. However, we have three major concerns about OMOP's recent attempts to achieve these goals.

Our first major concern is that most of the 'base truths' against which OMOP assessed pharmacoepidemiologic methods may not actually be true. To wit, 48 % of the 'positive control' drug–health outcome of interest (HOI) pairs examined by OMOP were based solely on case reports or case series [12]—hardly a sound basis for evaluating the validity of controlled epidemiologic studies. An additional 19 % of the positive controls had non-randomized epidemiologic studies as their basis [12]. Of course, validating a method against the results of studies that used the same method is circular. In contrast, by OMOP's count, only 32 % of the 'positive control' pairs were supported by randomized clinical trials [12]. Surely we need better reason to believe that 'base truths' against which research methods are evaluated are actually true.

Our second major concern is the potentially poor validity of three of the four HOI algorithms used by OMOP [13]. OMOP's studies examined acute myocardial infarction (AMI), acute kidney injury (AKI), acute liver injury (ALI), and gastrointestinal bleeding (GIB). OMOP's algorithm for GIB would probably perform well given that positive predictive values (PPVs; the proportion of putative cases identified by an algorithm that are actually true cases of that HOI) of 71 to 88 % have been found for similar algorithms [14–16]. However, OMOP's algorithm for AMI did not restrict cases to those for which the AMI diagnosis was in the principal position (i.e., the ostensible reason for the hospitalization), as is recommended [17], nor did it match an algorithm that has been widely used in recent years [17]. Given this, it is difficult to estimate the PPV of OMOP's primary AMI algorithm, although it might range from 52 to 95 % [17]. OMOP's primary algorithm for AKI would be expected to have a PPV ranging from 48 to 96 %, depending on the reference standard used to define a true case [18–21]. These variable PPVs for AKI highlight the inconsistent accuracy of diagnoses of renal conditions, which can lead to unacceptably low PPVs [22]. Identifying ALI from healthcare claims is even more challenging [23]. Some of the diagnostic codes used in OMOP's primary ALI algorithm have poor-to-moderate PPVs (range 7 to 54 %) [23]. In addition to chart-based validation studies cited above, OMOP investigators reported that their primary and alternate HOI algorithms had poor validity when evaluated against expert review of claims and laboratory

S. Hennessy · C. E. Leonard (✉)
Center for Pharmacoepidemiology Research and Training, Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
e-mail: celeonar@mail.med.upenn.edu

S. Hennessy
e-mail: hennessy@upenn.edu

values rather than medical records, with PPVs ranging from 1 to 56 % for AMI, 12 to 82 % for AKI, and 0 to 52 % for ALI [24]. It is unrealistic to expect to reproduce known associations using HOI algorithms with poor validity.

Our third major concern is the interpretation by Overhage et al. that their findings inform knowledge about the validity of the underlying epidemiologic designs themselves rather than (at best) the specific choices made in the implementation of those designs. There are sound theoretical bases for the epidemiologic research designs commonly used in pharmacoepidemiology, including cohort, case-control and self-controlled studies [25]. In the absence of bias (a key condition!), each design consistently yields a measure of association that has causal interpretability. Problematically implemented studies do not invalidate the underlying research designs, just those implementations.

Developing and evaluating methods to eliminate (or at least reduce) error is the methodologic imperative of pharmacoepidemiology. Within this imperative, a key methodologic challenge is the complete and accurate identification of important HOIs in large populations in whom drug exposure is recorded. This challenge can be addressed by devising and testing better approaches to identify HOIs. Another key methodologic challenge is the development and evaluation of approaches to mitigate confounding. Such approaches continue to be developed and applied in pharmacoepidemiology, and include self-controlled designs [26], instrumental variable approaches [27], the prior event rate ratio [28], high-dimensional propensity score methods [29], and statistical modeling approaches that accommodate time-varying confounding [30]. Each approach is applicable in different settings, and it is unrealistic to expect that any one method will be useful for all kinds of questions. Empiric evaluations of the performance of approaches to mitigate confounding are especially challenging, since we have little sense as to how faithfully our simulation studies reflect real-world confounding, which of course varies by study question. Therefore, carefully testing methods to mitigate confounding against thoughtfully selected known positive and negative associations using valid HOIs is therefore a useful and necessary exercise, and one that we hope to see much more of.

Funding No sources of funding were used to assist in the preparation of this letter.

Conflict of interest Sean Hennessy and Charles E. Leonard are collaborators in the Mini-Sentinel pilot project sponsored by the United States Food and Drug Administration (FDA) and coordinated by Harvard Pilgrim Health Care Institute (HPHCI). The views expressed herein are their own and may not represent the views of Mini-Sentinel, HPHCI, or FDA.

References

1. Overhage JM, Ryan PB, Schuemie MJ, Stang PE. Desideratum for evidence based epidemiology. Drug Saf. 2013;36(Suppl 1):S5–14.
2. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. Am J Epidemiol. 2006;163:783–9.
3. Amir Y, Sharon I. Replication research: a "must" for the scientific advancement of psychology. J Soc Behav Pers. 1990;5:51–69.
4. McCullough BD. Open access economics journals and the market for reproducible economic research. Econ Anal Policy. 2009;39:117–26.
5. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature. 2014;505:612–3.
6. Casadevall A, Fang FC. Reproducible science. Infect Immun. 2010;78:4972–5.
7. Janssen P, Enright AJ, Audit B, Cases I, Goldovsky L, Harte N, Kunin V, Ouzounis CA. COmplete GENome Tracking (COGENT): a flexible data environment for computational genomics. Bioinformatics. 2003;19:1451–2.
8. Peng RD. Reproducible research in computational science. Science. 2011;334:1226–7.
9. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539–58.
10. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ. 2001;323:42–6.
11. Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. II: replicate variability and comparison of studies that agree and disagree. Stat Med. 1987;6:733–44.
12. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Saf. 2013;36(Suppl 1):S33–47.
13. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31:4401–15.
14. Schelleman H, Bilker WB, Brensinger CM, Wan F, Yang YX, Hennessy S. Fibrate/statin initiation in warfarin users and gastrointestinal bleeding risk. Am J Med. 2010;123:151–7.
15. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology. 1996;7:101–4.
16. Cattaruzzi C, Troncon MG, Agostinis L, Garcia Rodriguez LA. Positive predictive value of ICD-9th codes for upper gastrointestinal bleeding and perforation in the Sistema Informativo Sanitario Regionale database. J Clin Epidemiol. 1999;52:499–502.
17. Metcalfe A, Neudam A, Forde S, Liu M, Drosler S, Quan H, Jette N. Case definitions for acute myocardial infarction in administrative databases and their impact on in-hospital mortality rates. Health Serv Res. 2013;48:290–318.
18. Waikar SS, Wald R, Chertow GM, Curhan GC, Winkelmayer WC, Liangos O, Sosa MA, Jaber BL. Validity of international classification of diseases, ninth revision, clinical modification codes for acute renal failure. J Am Soc Nephrol. 2006;17:1688–94.
19. Waikar SS, Curhan GC, Wald R, McCarthy EP, Chertow GM. Declining mortality in patients with acute renal failure, 1988 to 2002. J Am Soc Nephrol. 2006;17:1143–50.

20. Liangos O, Wald R, O'Bell JW, Price L, Pereira BJ, Jaber BL. Epidemiology and outcomes of acute renal failure in hospitalized patients: a national survey. Clin J Am Soc Nephrol. 2006;1:43–51.
21. Wald R, Liangos O, Waikar SS, Pereira BJG, Jaber BL. The identification of patients with acute renal failure (ARF) using administrative data: a validation study [abstract]. J Am Soc Nephrol. 2005;16:F-PO1031.
22. Vlasschaert ME, Bejaimal SA, Hackam DG, Quinn R, Cuerden MS, Oliver MJ, Iansavichus A, Sultan N, Mills A, Garg AX. Validity of administrative database coding for kidney disease: a systematic review. Am J Kidney Dis. 2011;57:29–43.
23. Lo Re V 3rd, Haynes K, Goldberg D, Forde KA, Carbonari DM, Leidl KB, Hennessy S, Reddy KR, Pawloski PA, Daniel GW, Cheetham TC, Iyer A, Coughlin KO, Toh S, Boudreau DM, Selvam N, Cooper WO, Selvan MS, VanWormer JJ, Avigan MI, Houstoun M, Zornberg GL, Racoosin JA, Shoaibi A. Validity of diagnostic codes to identify cases of severe acute liver injury in the US Food and Drug Administration's Mini-Sentinel Distributed Database. Pharmacoepidemiol Drug Saf. 2013;22:861–72.
24. Hansen RA, Gray MD, Fox BI, Hollingsworth JC, Gao J, Zeng P. How well do various health outcome definitions identify appropriate cases in observational studies? Drug Saf. 2013;36(Suppl 1):S27–32.
25. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
26. Maclure M, Fireman B, Nelson JC, Hua W, Shoaibi A, Paredes A, Madigan D. When should case-only designs be used for safety monitoring of medical products? Pharmacoepidemiol Drug Saf. 2012;21(Suppl 1):50–61.
27. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Instrumental variable analysis of secondary pharmacoepidemiologic data. Epidemiology. 2006;17:373–4.
28. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies. Pharmacoepidemiol Drug Saf. 2012;21(Suppl 2):60–8.
29. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009;20:512–22.
30. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11:550–60.