

Moving Along the Yellow Brick (Card) Road?

Stephen J. W. Evans

© Springer International Publishing Switzerland 2013

A recent issue of the Royal Statistical Society magazine “Significance” had an interesting article about the human tendency to be over-confident and the authors conclude “At the very least it is important for decision-makers to be aware that people are prone to overconfidence, and that to assume one is not is to unwittingly fall prey to the bias” [1]. From my experience of reviewing medical research articles, I find authors to be very over-confident of the strength of evidence provided by their research. This applies to randomised trials but especially to observational research.

In the same issue of Significance, on page 19 “Dr. Fisher” effectively notes this as well, though he describes his change in perspective when moving from author to referee. Being honest, I think it likely that I have been over-confident in my own research or opinion, but I like to think that in my mature years I have become more realistic both as author and as referee!

The OMOP research used data from Truven Health Analytics (formerly the Health Business of Thomson Reuters), and includes MarketScan[®] Research Databases, represented with MarketScan Lab Supplemental (MSLR, 1.2 m persons), MarketScan Medicare Supplemental Beneficiaries (MDCR, 4.6 m persons), MarketScan Multi-State Medicaid (MDCD, 10.8 m persons), MarketScan Commercial Claims and Encounters (CCAE, 46.5 m persons). Data also provided by Quintiles[®] Practice Research Database (formerly General Electric’s Electronic Health Record, 11.2 m persons) database. GE is an electronic health record database while the other four databases contain administrative claims data.

S. J. W. Evans (✉)

Department of Medical Statistics, Room 37b, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

e-mail: stephen.evans@lshtm.ac.uk

URL: <http://www.lshtm.ac.uk/people/evans.stephen>

OMOP is an empirically-based project to find good methods for detecting possible new adverse effects of medicines using databases from healthcare organisations. The US Congress has required that the FDA has available 100 million people’s data for post-marketing surveillance. This very idea may show over-confidence in believing that having the data available will mean that real effects will be detected reliably.

Overall the papers in this issue show clearly that there is considerable variation in the measures of association between drugs and adverse events. This is true both for those associations believed to be real adverse drug reactions and those believed to be coincidental. There are some problems in being sure of a gold standard, and this is acknowledged in these papers, but even with such issues it is clear that variability is much greater than is captured by a confidence interval or significance test. This has been well known for a long time and the excellent article by Maclure and Schneeweiss [2] sets out 11 domains that can lead to bias (and hence variability beyond sampling error). The first eight relate to the data and methods while the last three occur after the results are set out. Greenland suggests that such multiple biases can and should be modelled in a Bayesian framework [3].

The papers here are an empirical demonstration that variability in results in this context will occur, depending on:

1. The database used
2. The design
3. Parameters of the design such as “risk windows”
4. Statistical method used, though this is mainly related to the design

Some of the findings are not surprising therefore, but perhaps the magnitude of the variability is greater than many would expect.

It is also clear that what the authors describe as “self-controlled” methods tend to have a higher predictive accuracy than those that rely solely on between-patient comparisons. Again, this is not really surprising, since, where self-controlled methods’ assumptions are met, then their control of unmeasured but fixed (over time) confounders is always better than other methods. Perhaps this shows that the assumptions may not be as important as control of confounding.

What is also clear is that no single method with particular design parameters performs uniformly better than others. It also seems that different adverse events may need different approaches, again not really surprising when one considers the different pharmacology and biology involved.

What is remarkable about the OMOP project is the total transparency. The library of methods, the ability to exactly reproduce the results is, I think, unique in epidemiological or drug safety research. Critics, and there will be some, have the opportunity to show where there are errors in the results. However, the critics may have stronger grounds in questioning some of the interpretations.

Schneeweiss et al. [4] have shown that restriction in database studies may obtain more reliable, in the sense of being more similar to randomised trial, results. While there are some messages for pharmacoepidemiology in general, these may be more limited than the first paper in the series suggests. Maclure and Schneeweiss [2] note that their paper “can also be misused by pessimists who believe epidemiologic evidence is hopelessly biased”. In quoting Young the authors (Overhage et al. [5]) seem to align themselves with such pessimism. I cannot agree. Golder et al. [6] have shown that carefully conducted observational studies of adverse effects are more similar to randomised evidence than they expected. Rawlins [7] has also argued from a decision-maker’s perspective that epidemiology has a major contribution.

The messages for scanning databases without prior hypotheses (signal detection or generation), are more clear. It is not as easy as some of us supposed [8] the demise of spontaneous reporting (“yellow cards” in the UK) is not yet here. The overall conclusions of Gagne et al. [9] have not really been overturned, and the final conclusion of a key paper in this series (Ryan et al. [10]), comparing the performance of methods, is that “Observational healthcare data can inform risk identification of medical product effects on acute liver injury, acute myocardial infarction, acute renal failure and gastrointestinal bleeding.” This shows total pessimism and over-confidence are both wrong. In some senses a similar conclusion, that no single

method performs uniformly better and none is really excellent at distinguishing real from false effects, has been reached by a totally independent evaluation of some different methods in a UK General Practice Database [11]. We are not there yet, with the solution to problems of drug safety, but we are moving in the right direction.

Acknowledgements This article was published in a supplement sponsored by the Foundation for the National Institutes of Health (FNIH). The supplement was guest edited by Stephen J.W. Evans. It was peer reviewed by Olaf H. Klungel who received a small honorarium to cover out-of-pocket expenses. S.J.W.E has received travel funding from the FNIH to travel to the OMOP symposium and received a fee from FNIH for the review of a protocol for OMOP. O.H.K has received funding for the IMI-PROTECT project from the Innovative Medicines Initiative Joint Undertaking (<http://www.imi.europa.eu>) under Grant Agreement no 115004, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/2007–2013) and EFPIA companies’ in kind contribution.

References

1. Mannes A, Moore D. I know I’m right! A behavioural view of overconfidence. *Significance*; 2013, pp. 10–4.
2. Maclure M, Schneeweiss S. Causation of bias: the episcopo. *Epidemiology*. 2001;12:114–22.
3. Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc A*. 2005;168(2):267–306.
4. Schneeweiss S, Patrick AR, Stürmer T, Brookhart MA, Avorn J, Maclure M, Rothman KJ, Glynn RJ. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care*. 2007;45(10 Suppl 2):S131–42.
5. Overhage JM, Ryan PB, Schuemie MJ, Stang PE. Desideratum for Evidence Based Epidemiology. *Drug Saf* (in this supplement issue). doi:10.1007/s40264-013-0102-2.
6. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med*. 2011;8:e1001026.
7. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet*. 2008;372:2152–61.
8. Waller PC, Evans SJ. A model for the future conduct of pharmacovigilance. *Pharmacoepidemiol Drug Saf*. 2003;12:17–29.
9. Gagne JJ, Fireman B, Ryan PB, Maclure M, Gerhard T, Toh S, Rassen JA, Nelson JC, Schneeweiss S. Design considerations in an active medical product safety monitoring system. *Pharmacoepidemiol Drug Saf*. 2012;21(Suppl 1):32–40.
10. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, Reich C, Schuemie MJ, Madigan D. A Comparison of the Empirical Performance of Methods for a Risk Identification System. *Drug Saf* (in this supplement issue). doi:10.1007/s40264-013-0108-9.
11. Reps JM, Garibaldi JM, Aickelin U, Soria D, Gibson JE, Hubbard RB. Comparison of algorithms that detect drug side effects using electronic healthcare databases. *Soft Comput*. doi:10.1007/s00500-013-1097-4.