

Data mining in pharmacovigilance: lessons from phantom ships

Manfred Hauben · Lester Reich ·
Eugène P. Van Puijenbroek · Charles M. Gerrits ·
Vaishali K. Patadia

Received: 18 January 2006 / Accepted: 28 June 2006 / Published online: 3 August 2006
© Springer-Verlag 2006

Introduction

Pharmacovigilance experts devote considerable effort to post-marketing surveillance of adverse drug reactions (ADRs). Although the prepared mind of the pharmacovigi-

M. Hauben · L. Reich
Risk Management Strategy, Pfizer Inc,
New York, NY, USA

M. Hauben
Departments of Pharmacology and Community and Preventive
Medicine, New York Medical College,
Valhalla, NY, USA

M. Hauben
Department of Medicine, New York University,
School of Medicine,
New York, NY, USA

M. Hauben
School of Information Systems, Computing and Mathematics,
Brunel University,
West London, UK

E. P. Van Puijenbroek
Netherlands Pharmacovigilance Centre Lareb's-Hertogenbosch,
Groningen, The Netherlands

C. M. Gerrits
Department of Global Pharmacoepidemiology and Health
Outcomes Research, Takeda Global Research
and Development Inc,
Lincolnshire, IL, USA

V. K. Patadia (✉)
Pharmacoepidemiology, Global Drug Safety,
Amylin Pharmaceuticals,
San Diego, CA, USA
e-mail: vpatad1@yahoo.com

lance expert remains the cornerstone of this process [1], statistical algorithms, also known as data mining algorithms (DMAs), are being promoted as supplementary tools for safety reviewers. Opinions vary on their utility and optimum deployment mainly because their use has not been completely validated for various reasons, including a lack of consensus on gold standards for causality. True positive associations may be inherently more interesting, but constructing reference sets for validation also require identification of “true negatives” for measuring performance of DMAs.

Occasionally, drug-event associations (DEAs), originally considered credible based on traditional pharmacovigilance monitoring, are discounted with various levels of certitude after further investigation. We refer to these DEAs “phantom ships” [2]. Phantom associations may be discounted through epidemiological evidence, careful clinical analysis of the individual cases, and/or based on fundamental clinical pharmacological principles [3–9].

Objective

To highlight some previously ignored decision-theoretic aspects of signal detection using common implementation of two DMAs applied to eight potential “phantom” associations.

Methods

Two authors (M.H. and E.v.P.) selected a convenience sample of drug-event combinations (DECs), which could be

Table 1 SDRs of phantom associations based on a frequentist and an empirical Bayesian DMAs/metrics: number of reports to SDR and (year of first signal)

Drug-event combination number	Drug-MedDRA preferred term(s)	First year of publication [Publication reference number]	PRR>2, Chi sq>4, n> 2	EB ₀₅ >2	EBGM>2
DEC1	Celecoxib-cross sensitivity reaction	2001 [14]	3 (1999)	–	3 (1999)
DEC2	Diazepam-cleft lip/cleft palate	1974 [15]	3 (1986)	–	2 (1976)
DEC3	Doxylamine-multiple congenital abnormalities	1976 [16]	5 (1983)	5 (1983)	2 (1970)
DEC4	Levodopa-malignant melanoma	1974 [17]	4 (1997)	3 (1996)	2 (1993)
DEC5	Metronidazole-alcohol intolerance	1980 [18]	3 (1993)	8 (1997)	3 (1993)
DEC6	Nicotine-myocardial infarction	1993 [19]	3 (1984)	93 (1992)	93 (1992)
DEC7	Reserpine-breast neoplasm	1974 [20]	3 (1976)	4 (1977)	2 (1975)
DEC8	Simvastatin-cataract	1989 [21]	6 (1993)	19 (1996)	6 (1993)

identified as ‘phantom DEC.’ These are listed in Table 1. Four currently used metrics from two types of disproportionality analysis, a frequentist method (i.e. standard PRRs [10]), and an empirical Bayesian method (i.e. stratified MGPS [11]) were applied to the FDA-AERS database through the 3Q2003¹. For each metric/threshold, the timing of the first statistical disproportionality—hereafter referred to as a signal of disproportionate reporting (SDR) [12]—was identified. A MEDLINE database search was used to identify the first literature citation for each of the DEC. For PRRs, an SDR was defined as a PRR >2 and Chi sq>4 and case count >2 [10]. For MGPS, we used the commonly-cited threshold of EB₀₅>2, N>0 [13] and an additional threshold EBGM>2, N>0.

Results

Both frequentist and empirical Bayesian algorithms were associated with SDRs for all of the associations. All generated an SDR for all phantom associations with the exception of the commonly cited MGPS metric EB₀₅>2 for DEC 1 and 2. Literature reports preceded an SDR in five instances with both DMAs (see Table 1).

Discussion and conclusions

Both DMAs generated SDRs for all selected phantom associations for one or more metrics. For DEC 1 and 2, EB₀₅>2 was the only threshold metric that discriminated such phenomena. This is not surprising because it may be the most “severe” of the metrics, in that it incorporates *empirical* Bayesian shrinkage plus an additional frequentist element of shrinkage due to the use of the lower bound of

the 95% posterior interval. While we were unable to review every case of each association to determine the quality of the clinical evidence, our sample included published case reviews that were notable for a lack of evidence to support an association.

Defining a misclassification error when evaluating DMAs has been the subject of vigorous debate. Regarding false positive misclassification, some argue that if the data was misleading, but the DMA accomplished its intended objective of identifying associations not obviously identifiable at the outset as spurious (i.e. warranting further investigation), then it should not be counted as a misclassification by the DMA. Another view based on the interest in the incremental utility of DMAs versus traditional approaches, is that such scenarios represent misclassification by *both* traditional and computational approaches. Although traditional and computational approaches to signal detection have distinctive and complementary features [22], a corollary lesson is that since they are both related by the same dataset, they share common properties so their misclassifications errors are likely to be correlated.

Although classification errors are to be expected with any screening tool, the results of the present study constitute a further caution against “seduction bias”—the tendency to over-interpret findings generated from algorithms with an extensive mathematical framework, when they are susceptible to many of the same reporting biases and artifacts as traditional approaches [22, 23]. There are a myriad of factors [24–30] that influence reporting (e.g., attention of medical and/or lay press) and which therefore result in misclassification by both traditional and computational methods. Literature reports preceded an SDR in five instances with both DMAs (Table 1). Hence, previous publications in the literature may be a predisposing factor for yielding statistical associations when data mining FDA-AERS database.

It is also especially noteworthy that most of the selected phantom associations were highlighted based upon small

¹Using WebVDME 4.0 by Lincoln Technologies (Waltham, MA)

numbers of reports (Table 1). Often, the ADRs involved in such ‘phantom ships’ associations constitute signs or symptoms, which have low background incidence rates and are rarely reported ADRs for other drugs, and therefore small numbers of the association are sufficient to yield a statistically significant effect when applying DMAs.

What is the significance of the greater discriminatory behavior of the $EB_{05} > 2$ threshold in this exercise? Some investigators assign priority to the “less is more” principle—namely that a metric is superior if it presents the user with fewer potential associations for evaluation. Notwithstanding the findings from our small and non-systematic sample, this remains only opinion at this time since there is no clear decision theoretic framework to guide such assessment, [31] and the relative importance of sensitivity versus specificity may be situation dependent.

Previous publications have not fully explored these issues and some answers are accepted before all the questions have even been formulated. For example, what are the relative benefits and opportunity costs of earlier detection of both true and spurious associations? Earlier detection with a smaller number of cases is always assumed to be advantageous. But if the association cannot be clarified until additional cases are submitted, and this coincides with initial detection by a less sensitive method, then earlier detection by the more sensitive method merely imposed an additional burden of monitoring over time without earlier resolution, akin to lead-time bias in medical screening. Conversely, earlier detection may allow more timely implementation of highly focused and intensified follow-up data capture procedures which itself could lead to earlier resolution. Analogous considerations could apply to spurious associations. A more careful and systematic analysis of the utilities and costs associated with the use of DMAs in real-world pharmacovigilance scenarios could yield added benefits and insights over the usual published data mining exercises [31].

We believe that certain phantom ships might be included within a larger reference set for understanding performance of DMAs relative to traditional approaches. Although many questions remain about the optimal approach to such validation exercises [32], human interpretation of the results remains pivotal [23].

Acknowledgement We would like to thank Barbara J. Stephenson RN, MSC (Epi & Biostats) for her help in critical review of this article.

References

- Trontell A (2004) Expecting the unexpected-drug safety, pharmacovigilance, and the prepared mind. *N Engl J Med* 351:1385–1387
- Stricker BH (2002) Pharmacovigilance: a case of phantom ships and Russian roulette. *Ned Tijdsch Geneeskunde* 146:1258–1261
- Sober AJ, Wick MM (1978) Levodopa therapy and malignant melanoma. *JAMA* 240:554–555
- Fiala KH, Whetteckey J, Manyam BV (2003) Malignant melanoma and levodopa in Parkinson’s disease: causality or coincidence? *Parkinsonism Relat Disorders* 9:321–327
- Williams CS, Woodcock KR (2000) Do ethanol and metonidazole interact to produce a disulfiram-like reaction? *Ann Pharmacother* 34:255–257
- Siple JF, Schneider DC, Wanlass WA, Rosenblatt BK (2000) Levodopa therapy and risk of malignant melanoma. *Ann Pharmacother* 34:382–385
- Kleinhans M, Schmid-Grendelmeier PS, Burg G (1996) Levodopa und malignes melanoma - fallbericht und literaturübersicht ein beitrage zur frage des kausalzusammenhanges zwischen levodopa und der entwicklung eines malignen melanoms. *Del Hautgarzt* 47:432–437
- Toler S, Rodriguez I (2004) Not all sulfa drugs are created equal. *Ann Pharmacol* 38:2166–2167
- Johnson KK, Green DL, Rife JP, Limon L (2005) Sulfonamide cross-reactivity: fact or fiction? *Ann Pharmacol* 39:290–301
- Evans SJ, Waller PC, Davis S (2001) Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 10:483–486
- Dumouchel W (1999) Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 53(3):170–190
- Hauben M, Reich L (2005) Communication of findings in pharmacovigilance: use of term “signal” and the need for precision in its use. *Eur J Clin Pharmacol* 61(5–6):479–480
- Szarfman A, Machado SG, O’Neill RT (2002) Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA’s spontaneous reports database. *Drug Saf* 25(6):381–392
- Knowles S, Shapiro L, Shear NH (2001) Should celecoxib be contraindicated in patients who are allergic to sulfonamides? Revisiting the meaning of ‘sulfa’ allergy. *Drug Saf* 24:239–247
- Walker Be, Patterson A (1974) Induction of cleft palate in mice by tranquilizers and barbiturates. *Teratology* 10:159–163
- Miklovich L, Van den Berg BJ (1976) An evaluation of the teratogenicity of certain antiemetic drugs. *Am J Obstet Gynecol* 125:244–248
- Happle (1974) Malignant melanoma and L-dopa. Review of literature on the problem of causal relationship. *Fortschr Med* 92:1065
- Finegold SM (1980) Metronidazole. *Ann Intern Med* 93:585–587
- Dacosta A, Guy JM, Tardy B, Gonthier R, Denis L, Lamaud M, Cerisier A, Verneyre H (1993) Myocardial infarction and nicotine patch: a contributing or causative factor? *Eur Heart J* 14:1709–1711
- Anonymous (1974) Reserpine and breast cancer. *Lancet* 2: 669–671
- Behrens-Baumann W, Morawietz A, Thiery J, Creutzfeldt C, Seidel D (1989) Ocular side effects of the lipid-lowering drug simvastatin? A one year follow-up. *Lens Eye Toxic Res* 6:331–337
- Hauben M, Madigan D, Gerrits CM, Walsh L, Van Puijenbroek EP (2005) The role of data mining in pharmacovigilance. *Expert Opin Drug Saf* 4:929–948
- Hauben M, Patadia V, Gerrits C, Walsh L, Reich L (2005) data mining in pharmacovigilance: the need for a balanced perspective. *Drug Saf* 10:835–842
- Bateman DN, Sanders GL, Rawlins MD (1992) Attitudes to adverse drug reaction reporting in the Northern Region. *Br J Clin Pharmacol* 34:421–426
- Belton KJ (1997) Attitude survey of adverse drug-reaction reporting by health care professionals across the European Union.

- The European Pharmacovigilance Research Group. *Eur J Clin Pharmacol* 52:423–427
26. Belton KJ, Lewis SC, Payne S, Rawlins MD, Wood SM (1995) Attitudinal survey of adverse drug reaction reporting by medical practitioners in the United Kingdom. *Br J Clin Pharmacol* 39:223–226
 27. Cosentino M, Leoni O, Banfi F, Lecchini S, Frigo G (1997) Attitudes to adverse drug reaction reporting by medical practitioners in a Northern Italian district. *Pharmacol Res* 35:85–88
 28. Eland IA, Belton KJ, Van Grootheest AC, Meiners AP, Rawlins MD, Stricker BH (1999) Attitudinal survey of voluntary reporting of adverse drug reactions. *Br J Clin Pharmacol* 48: 23–637
 29. Williams D, Feely J (1999) Underreporting of adverse drug reactions: attitudes of Irish doctors. *Ir J Med Sci* 168:257–261
 30. De Bruin MI, Van Puijenbroek EP, Egberts AC, Hoes AW, Leufkens HG (2002) Non-sedating antihistamine drugs and cardiac arrhythmias – biased risk estimates from spontaneous reporting systems? *Br J Clin Pharmacol* 53:370–374
 31. Chan KA, Hauben M (2005) Signal detection in pharmacovigilance: empirical evaluation of data mining tools. *Pharmacoepidemiol Drug Safe* 14:597–599
 32. Valenstein PN (1990) Evaluating diagnostic tests with imperfect gold standards. *Am J Clin Pathol* 93:252–258