CONFERENCE PAPER

Drug Safety 2007; 30 (7): 627-630
0114-5916/07/0007-0627/$44.95/0

Signal Detection in the Pharmaceutical Industry

Integrating Clinical and Computational Approaches

Manfred Hauben[1,2,3,4]

1 Risk Management Strategy, Pfizer Inc., New York, New York, USA
2 Department of Medicine, New York University School of Medicine, New York, New York, USA
3 Department of Community and Preventative Medicine, New York Medical College, Valhalla, New York, USA
4 School of Information Systems, Computing and Mathematics, Brunel University, London, UK

Abstract

Drug safety profiles are dynamic and established over time using multiple, complimentary datasets and tools. The principal concern of pharmacovigilance is the detection of adverse drug reactions that are novel by virtue of their clinical nature, severity and/or frequency as soon as possible with minimum patient exposure. A key step in the process is the detection of 'signals' that direct safety reviewers to associations that might be worthy of further investigation. Although the 'prepared mind' remains the cornerstone of signal detection safety reviewers seeking potential signals by scrutinising very large, sparse databases may find themselves 'drowning in data but thirsty for knowledge'. Understandably, health authorities, pharmaceutical companies and academic centres are developing, testing and/or deploying computer-assisted database screening tools (also known as data-mining algorithms [DMAs]) to assist human reviewers. The most commonly used DMAs involve disproportionality analysis that project high-dimensional data onto two-dimensional (2×2) contingency tables in the context of an independence model. The objective of this paper is to extend the discussion of the evaluation, potential utility and limitations of the commonly used DMAs by providing a 'holistic' perspective on their use as one component of a comprehensive suite of signal detection strategies incorporating clinical and statistical approaches to signal detection – a marriage of technology and the 'prepared mind'. Data-mining exercises involving spontaneous reports submitted to the US FDA will be used for illustration. Potential pitfalls and obstacles to the acceptance and implementation of data mining will be considered and suggestions for future research will be offered.

"*Our initial interest in data mining algorithms is in what they can tell us that we do not already know.*"

Implementing a signal detection methodology is a multidisciplinary effort, synthesising various approaches, from the statistician to the clinical pharmacologist, for example.

In the beginning, there was possibly some overenthusiastic publication bias around data-mining algorithms (DMAs). In addition, many imprecise statements were made in various forums, e.g. data mining is useful, data mining should never be used in isolation, data mining should never replace clinical judgement, and these statements can have multiple interpretations.

Overall, one may say that there are different 'styles' of signal detection. Clinical reasoning versus computational approach is not a dichotomous choice, as many signal detection programmes fall along a continuum representing a mix of the two perspectives.

There is considerable semantic variability associated with the use of the term 'signal' and a statistical calculation based on spontaneous reports does not necessarily rise to the level of a signal when considered in a biological vacuum. A signal should incorporate some clinical scientific context, and signal detection is not necessarily a point but a process that gradually emerges over time. Signal of disproportional reporting (SDR) is the term coined to refer to statistical calculation alone devoid of clinical context to emphasise that such numerical results reflect reporting tendencies that could be a function of numerous factors unrelated to causality such as statistical noise confounding and/or reporting artifacts.

There are two main uses for the DMAs. In addition to hypothesis generation, another relevant contribution is in refining an index of suspicion for a signal detected by other means.

For purposes of signal detection, it is very important to define what is to be expected from a DMA before its implementation, e.g. to detect adverse events that would otherwise remain undetected, to detect the same adverse events but at earlier time points, to detect the same adverse events at the same time but more efficiently, which means spending less time evaluating false leads or with less person-time resources or to provide a safety net for human cognitive or system errors.

In terms of hypothesis generation, events can be conceptualised into three broad categories. The designated medical events: these are serious, rare, with high drug-attributable risk and high and asymmetric cost of misclassification. At the other end, there are the events that are insignificant in terms of public health and of limited importance as long as a sensible triage approach is in place. Then there are all the other events that fall in between, which include various types of event such as paradoxical reactions and 'surprise reactions'. Understanding the 'sample space' of adverse events may help us formulate an overall pharmacovigilance logic.

In terms of the role to refine and index a suspicion, we can consider three types of clinical signals. There are clinical signals that are very compelling, with cases providing cogent clinical evidence of a strong signal. Statistical calculations do not have a great deal of incremental value in this situation, being either of marginal additional value or actually misleading. Similarly, there are cases that are compelling non-signals, for example a combination of case reports that are totally spurious upon review and where a statistical disproportionality may have originated from biased reporting. The clinical data suggest that the statistical information is more likely to be misleading and data mining has less value. Thirdly, a common scenario is that of clinically ambiguous signals. In those cases, the clinical information does not point clearly in one direction and it is very difficult to reach a conclusion and the statisti-

cal calculations may be one useful piece of the puzzle to add to the body of evidence, though by no means definitive by itself. To sum up, when the results are discordant, cogent clinical evidence tends to trump statistical arithmetic based on flawed data.

When carrying out data mining, a huge number of choices are presented to the analyst, such as choosing the suspect drug versus the suspect plus concomitant drugs, the algorithm metric/threshold, stratified versus unstratified analysis, data base, the level of the dictionary hierarchy or the use of standardised MedDRA (Medical Dictionary for Regulatory Activities) queries, to name a few. There is therefore a huge number of analysis each defined by the specific selections. This is a double-edged sword, maximising exploratory capacity but also making data-mining exercises susceptible to confirmation bias in which multiple analysis are tried but one selected based on the fit of the results to pre-existing expectations.

One key question is what data mining can tell us that we do not already know, or in other words, what new information they can offer that we would not have had otherwise.

At Pfizer, road testing the common implementations of these methods was found to be immensely useful. US FDA AERS (Adverse Event Reporting System) data were the source mostly used for data mining using the proportional reporting ratio and multi-item gamma poisson shrinker (MGPS). Commonly cited disproportionality thresholds were employed in retrospective two-dimensional analyses at the MedDRA Preferred Term level with diverse pharmacovigilance scenarios. The general database was used as the background.

1. Lessons Learned

A prepared mind is very critical. The way we do signal detection is very clinically front-loaded, retaining clinical judgment and reasoning from the very first step.

It is hard to feel comfortable relying totally on the DMA. They may miss relevant associations, generate 'false positives', and there are rate limiting problems with the data. In this context, the phrase 'automated signal detection' is not optimal since signal detection should not be a truly automated procedure devoid of input and interpretation of the pharmacovigilance domain expert.

Both 'false positives' and 'false negatives' are significant challenges and can have major impacts. Finding an appropriate balance between 'sensitivity' and 'specificity' is vital for any signal detection tool.

The MedDRA hyper-granular dictionary can be challenging when making comparisons between different algorithm results. One has to be equally careful and impartial when establishing case definitions.

'Paradoxical reactions' constitute a type of event requiring a certain level of case-by-case review. Seizures with anti-seizure medication, anaphylaxis with corticosteroids, paradoxical hypertension with antihypertensives, are some examples. They can be very allusive but disregarding them may lead to missing important safety issues.

One area where data mining may be very useful is in detecting 'surprise reactions'. These may be prone to being discounted on manual review of adverse event lists. They are not serious enough to be designated medical events and they are not clearly related to the drugs' clinical pharmacology. One example of this type of association is oligohidrosis with topiramate.[1]

A real breakthrough would be to link knowledge mining to data mining so that the safety reviewer could readily access the nuances of the suspected drug's pharmacology.

'Seduction bias' occurs in signal detection when sophisticated statistical terminology and dazzling graphical presentations make people believe that DMAs can tell you more than they really do.

2. Implementation

The Pfizer process incorporates the DMAs with some triage logic and included the age of the drug as one element. Having a 'worst first' list of sentinel events, for which sensitivity is emphasised, provides more flexibility for deploying quantitative tools. The process does not rely on the DMAs for detecting designated medical events. Algorithms are used in parallel with clinical methods, not in series.

For an old drug, presumed to be well known, a higher threshold may be defined since most of the findings are likely to be false positives and true safety issues would show a persistent signal. Keep in mind that a 'worst first' list could still be applied to old drugs.

Bayesian methods are statistically more elegant. However, when doing signal detection front-loaded with clinical reasoning, all algorithms are workable tools with small differences between them. This is in agreement with one group of researchers at the FDA guidance who compared frequentist with empirical Bayesian metrics used in VAERS (the Vaccine Adverse Event Reporting System) and who concluded that the statistical differences may not always translate into clinically significant differences.[2]

Data mining will probably help detect relevant safety issues but as currently used they are unlikely to detect large numbers of previously unknown safety issues of major significance. Its major value may be in helping to refine an index of suspicion and to define which direction to go forward. The bottom line, improving medical cognition must not be neglected and using DMAs as a substitute for clinical knowledge and judgment does not serve the interests of patient safety.

Acknowledgements

No sources of funding were used to assist in the preparation of this paper. The author has no conflicts of interest that are directly relevant to the content of this paper.

References

1. Arcas J, Ferrer T, Roche MC, et al. Hypohidrosis related to the administration of topiramate to children. Epilepsia 2001; 42 (10): 1363-5
2. Banks D, Woo EJ, Burwen DR, et al. Comparing mining methods on the VAERS database. Pharmacoepidemiol Drug Saf 2005; 14 (9): 601-9

Correspondence: Dr *Manfred Hauben*, Pfizer Inc, 235 E. 42nd St., New York, NY 10017-5755, USA.
E-mail: Manfred.Hauben@Pfizer.com