Drug Safety 2008; 31 (3): 271-272
0114-5916/08/0003-0271/$48.00/0

The 'Power' of Signal-Detection Algorithms

In recent years, interest in the use of data mining in pharmacovigilance has increased greatly.[1] This is mainly a result of the availability of commercial signal-detection algorithms that can be easily employed to search huge spontaneous reporting system databases. These data-mining algorithms can search a database for reporting association based on statistical dependencies between drugs and adverse events (AEs) in excess of what would be expected if the drug and AE were independently distributed in the database of interest.

While there has been much discussion about sensitivity, specificity and predictive value of various signalling methods, there is no universally agreed definition of a 'true positive'. This lack of a gold standard has hampered efforts to address these issues in a systematic matter. The proposed use of surrogates such as labelling changes or appearance in the scientific literature is also associated with limitations that include regulatory and legal factors on product labelling and publication bias.

However, the abovementioned caveat notwithstanding, there remain very valid concerns that the use of computer-assisted, signal-detection algorithms on large databases may generate a very high volume of signal scores exceeding a predetermined level (statistical threshold). Many of these will represent AE/drug combinations that are not relevant to the understanding of potential drug safety issues. The observed elevated signal scores are attributable to other reasons such as coding issues, stimulated reporting, etc. These are commonly referred to as false-positive signals and represent the classic statistical Type I error.

A large volume of false-positive signals can put undue pressure on pharmacovigilance resources. Because of this concern, various methods that minimize the chance of detecting these false positives AE/drug combination have been employed. Several authors have suggested adjustments of the signalling criteria to reduce the volume of elevated signal scores. However, components of the statistical modelling and modifications that reduce the volume of elevated signals scores may also degrade the ability of the signal-detection algorithm to detect true-positive signals.

In a recent paper, Matsushita et al.[2] presented results from a Monte Carlo simulation study. The criteria for signal detection used in this study were as follows:

PRR > 2, Chi-square >4

IC −2SD >0

EB05 >2

for PRR, BCPNN and MGPS, respectively, where IC is the information component and EB05 is the lower fifth percentile of the posterior distribution.

The study found that the sensitivities of PRR, BCPNN and MGPS methods were unacceptably low (49%, 45% and 26%, respectively), whereas their specificities were reasonable (95%, 99.6% and 99.99%, respectively). This finding is not surprising at all. These commonly used criteria for signal detection are derived with an emphasis on the control of Type I error (specificity) and not Type II error (sensitivity).

None of the commonly used signalling criteria are suitable for directly controlling the sensitivity of the signal-detection algorithms. Note that, to increase sensitivity we need to reduce false negative signals (Type II error). In statistical term, reducing Type II error (β) is the same as increasing power ($1-\beta$) of a test criterion.

However, on a theoretical basis, we should strongly consider whether the debate concerning the adjustments of signal scoring to control for Type I error is appropriate when evaluating safety issues. Additionally, it is important to consider, given the heterogeneity of many large databases, the need to

establish different signalling thresholds for different AE/drug combinations based on the importance of that AE to public health concerns.

The issue at hand can be clearly demonstrated by a simple example. Assume that the background reporting rate of an AE of interest is 2.5 per 1000. Now if a pharmaceutical manufacturer is interested in monitoring a newly marketed drug for this specific AE (using the commonly cited signalling criterion PRR >2, chi-squared > 4, N ≥3), a signal will be detected if out of 2000 AE reports there are ≥11 reports of this specific AE of interest. A power calculation will show that a signalling threshold of 11 out of 2000 would only have about 42% chance to detect an expected PRR >2.

On the other hand, if the manufacturer wants to rule out an expected PRR >2, with at least 80% certainty (power), then the signalling threshold would be 7 out of 2000. This approach has an added advantage because the expected PRR and/or the sensitivity (power) for the signalling threshold can be manipulated directly for different possible drug/AE combinations (according to their public health implication). For example, if the manufacturer wants to rule out an expected PRR >1.5, with at least 90% certainty (power), then the signalling threshold would be 4 out of 2000.

In conclusion, a signalling threshold based on Type II error (power) should be considered for signal-detection methodology and may be preferred over that based on Type I error (specificity) when monitoring a newly marketed drug or for a medically important AE from a public health perspective. A likely consequence of this would be an increase in false-positive signals, which could be better managed through signal triaging. We intend to further explore this concept by performing additional analyses on theoretical databases of varying sizes and AE/drug combinations.

M. Rezaul Karim, *Stephen L. Klincewicz* and *Chuen L. Yee*
Benefit Risk Management, Johnson & Johnson Pharmaceutical Research & Development, Horsham, Pennsylvania, USA

Acknowledgements

The authors have no conflicts of interest directly relevant to the content of this letter.

References

1. Almenoff J, Tonning JM, Gould AL, et al. Perspectives on the use of data mining in pharmacovigilance. Drug Saf 2005; 28 (11): 981-1007
2. Matsushita Y, Kuroda Y, Niwa S, et al. Criteria revision and performance comparison of three methods of signal detection applied to the spontaneous reporting database of a pharmaceutical manufacturer. Drug Saf 2007; 30 (8): 715-26