ABSTRACTS

1 Data Mining in Pharmacovigilance Practice: Which Algorithm to Use?

Yang X,[1] *Freeman J,*[1] *Bwire R*[1]
(1)Global Drug Safety, Celgene Corporation, Summit, New Jersey, USA

Introduction: Data mining techniques are used in identifying safety signals in pharmacovigilance practice. Commonly used algorithms include proportional reporting ratio (PRR), reporting odds ratio (ROR) and multi-item gamma Poisson shrinkage (MGPS). Comparison of performance of these algorithms is a subject of ongoing discussion. There is no consensus on a single algorithm, which has resulted in use of multiple algorithms in practice. This is likely to create difficulties in interpreting results and increases the workload. We assessed the performance of PRR, ROR and MGPS in identifying safety signals.

Methods: A commercially available FDA AERS database through the second quarter of 2007 was analyzed. Signal scores for suspected drugs in the database were computed. Signals were identified by applying the commonly cited thresholds (PRR: N>2, PRR > 2 and $\chi^2 > 4$; ROR: N>2, $ROR_{05} > 1$ and ROR >2; MGPS: N > 2 and $EB_{05} > 2$). The signals identified by the algorithm under study were stored as a "primary" signal dataset. Then thresholds from the remaining 2 algorithms were applied separately on the "primary" signal dataset to identify the number of signals that would be returned by each of the 2 algorithms. All three algorithms were separately used to create a "primary" signal dataset.

Results: There were 154,158 signals identified using the PRR threshold, of which 99.98% were flagged as signals by using ROR and 31.6% by using MGPS. Of the 162,744 signals identified by ROR, 94.7% were also flagged by PRR and 30.0% by MGPS. The MGPS algorithm generated 48,766 signals with 100% of these signals also flagged by PRR and 99.98% flagged by ROR. For the 0.02% (27) signals detected by PRR but not by ROR and the 5.3% (8613) signals detected by ROR but not PRR, 96% of them had counts smaller than 10.

Conclusion: These results show that if the PRR or ROR is used it is not necessary to further assess signals using MGPS. PRR and ROR identify 100% and 99.98% of the MGPS signals respectively. Also it is not necessary to analyze signals with both PRR and ROR since there is an approximately 95% overlap in signals from both algorithms. Using either PRR or ROR suffices to identify signals for further analysis. The MGPS only identifies about 30% of PRR and ROR signals. The selection of MGPS could cut down the signal assessment workload by about 70%, however, there is a risk of missing potential signals.

Conflicts of interest: None declared