Necessity of Artificial Intelligence to Detect Signal in the Field of Pharmacovigilance

Yalamanchili Jahnavi[1*], Kevin James[2], Prizvan Lawrence Dsouza[3], Basavaraj B.V.[4] and Viswam Subeesh[5]

*Contact Author e-mail: jahnaviy.yalamanchili@gmail.com,

Contributors:
[*1, 2, 3] 5th Pharm- D Students,
[4]Associate Professor,
Department of Pharmaceutics
[5]Assistant Professor,
Department of Pharmacy Practice,
M.S.Ramaiah University of Applied Sciences,
Gnanagangothri Campus,
New BEL Road, M S R Nagar,
Bangalore, Karnataka,
INDIA - 560 054

Abstract

Determining the benefits and possible risks of the medications helps in the rational use of medications. Documentation of the adverse events or possible risks of the medication through spontaneous reporting as a part of post-marketing safety surveillance plays a pivotal role in the pharmacovigilance program. Signal detection of previously unknown adverse events can be identified through the data mining process of big data available in large databases. These databases provide individual case safety reports that can be analyzed using statistical methods such as qualitative and quantitative approaches to measure the association between drug-event combinations, thereby confirming a signal. These approaches are time-consuming and have their own limitations such as the possibility of false safety signals due to insufficient information, over-reporting, presence of confounding factors and variation in sample size. To overcome these limitations, artificial intelligence can be essential in the early detection of signals by reducing the human efforts and time for analysis. Further advancement of quantitative methods alongside artificial intelligence can improve the process of detection of true signals.

Key Words*: Artificial Intelligence, Signal Detection, Pharmacovigilance*

1. Introduction

Rational use of drugs is a process of overweighing the expected benefits over the possible risks. The adverse events or the possible risks of the drug after marketing authorization are more evident with the help of routine post-marketing safety surveillance which plays a major role in pharmacovigilance process[1]. Pharmacovigilance (PV) or Drug Safety is "A process that deals with the science and activities linking to the collection, detection, assessment, monitoring, and prevention of adverse effects or any other drug related problem". Pharmacovigilance is promoted by the World Health Organization (WHO) in order to ensure and enhance care and safety to patients concerning with the use of any pharmaceutical products[2].

Signal detection or signal generation is a crucial part of the pharmacovigilance process for many years. The ultimate aim of signal detection is to identify previously unknown Adverse Drug Reactions (ADRs) and provide guidance to reduce the risk of using that particular drug in the population[3, 6].

Signal as per WHO is defined as "Reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously". The signal generation process often requires more than one reported Information to show a relationship between the adverse event and the drug associated with it[4, 12]. Detection of signals generated mainly depends on the data mining process of huge spontaneous reporting systems that focuses on the detection of unknown ADRs[2]. Data mining of pharmacovigilance databases owned by pharmaceutical companies, drug regulatory authorities, or large healthcare providers, is a widely accepted approach due to the availability of big data and economical computing resources. These databases are easily accessible and therefore, Individual Case Safety Reports (ICSRs) can be retrieved and arranged into a well-organized format, whereby applying proper statistical methods give the statistical measures of association between a drug and the given ADR, thereby confirming the signal[4, 5, 6]

Even though various qualitative and quantitative methods are available for detection of signal, challenges such as accuracy and early detection of evidence-based signal still remains. Another important challenge is to reduce the amount of false-positive signals, which requires greater individual support for analysis. Thus, the application of artificial intelligence (AI) in

Pharmacovigilance signal detection process is essential to overcome the limitation of known quantification methods[7, 8, 9].

2. Methods of Data Mining

To detect a signal, qualitative methods as well as quantitative methods can be applied. Signal detection using qualitative method requires reviewing of the spontaneous ICSRs and scientific literature. This process is time-consuming and can only be applied to analyze a signal when the number of reported cases is small. Thus, quantitative methods act as an alternative to generate a signal for large data by performing evaluations through automated algorithms known as data mining algorithms. The methods involved are known as Disproportionality Analysis (DPA) which follows the assumption that the disproportionately higher Drug-Event Combination (DECs) in the database may act as new and important signal upon the removal of background noise. Frequentist and Bayesian approaches are two methods in DPAs[10, 11,12,13,14]. The Frequentist approaches requires the 2x2 table as shown in table 1 and are relatively easier to understand, compute and analyze[4].

Table 1. The 2x2 contingency table for computing Frequentist method

	Suspected drug	**All other drugs in the database**	**Total**
Suspected ADR	M	N	M + N
All other database ADR	P	Q	P + Q
Total	M + P	N + Q	M + N + P + Q

M - Number of reports that contain the suspected drug or suspected pharmaceutical product with the suspected ADR.

N - Number of reports that contain other drugs or other pharmaceutical product with the suspected ADR.

P – Number of reports that contain the suspected drug or suspected pharmaceutical product with other ADRs.

Q – Number of reports that contain other drugs or other pharmaceutical product with other ADRs.

3. Limitations of Signal Detection Methods and Spontaneous Data Utilized

- One of the major limitations of Spontaneous Reporting System (SRS) is associated with the methods used, such as Proportional Reporting Ratio (PRR), Relative Reporting Ratio (RRR) and Reporting Odds Ratio (ROR) which can give false safety signals due to variation in the sample size.
- Even though, the Spontaneous Reporting System (SRS) database represents a majority of DECs, the quantitative methods cannot be used for low count DECs due to assumed threshold criteria.
- The quality and bias of spontaneous data reports mainly include over-reporting of same incident, under-reporting of new cases, incorrect or incomplete information and incorrect association between drugs and adverse events.
- These databases also lack information regarding the population using the drug without an ADR association.
- The chances of getting false positive signals are high as the quantitative methods fails to correct the statistical thresholds intended for the quality and confounding factors, which includes patient medical history, comorbid conditions, concomitant drugs used and other underlying disease conditions. Social habits like smoking and alcohol consumption can also influence the results by acting as confounding factors.
- Unreliable data is also obtained due to the instability in the frequentist signal detection methods whereas the latest methods such as MGPS and BCPNN have got decreased sensitivity as they mainly focus on shrinking of the Relative Reporting Ratio or IC for sampling variance when the availability of the data regarding the drug-adverse event is less. Thus, these methods are not effective in detecting a signal for lesser exposed drugs to thePopulation and the drawback due to presence of confounding factors still persists[7, 8, 10, 15].

Table 2. Provides an insight regarding the computation and published threshold criteria of the two approaches with their advantages and limitations[1, 3, 4, 5,9,11].

χ^2= chi-square, n=reported cases number, CI=Confidence Interval, EBGM=Empirical Bayesian Geometric Mean, IC= Information Component, SD=Standard Deviation.

DPA	COMPUATION	PUBLISHED THRESHOLD CRITERIA	EXPLANATION
Frequentist Methods			
Proportional Reporting Ratio (PRR)	$\frac{M/(M+N)}{P/(P+Q)}$ $95\%CI = e^{\ln(PRR)\pm 1.96\sqrt{\frac{1}{M}+\frac{1}{N}+\frac{1}{P}+\frac{1}{Q}}}$	PRR ≥ 2 $\chi^2 \geq 4$ n ≥ 3	PRR is used to calculate the reporting rate of the suspected ADR of interest among all ADRs for the suspected drug of interest, compared to the reporting rate for all the drugs included in the database. Application and interpretation of PRR is easier and the method is more sensitive than Bayesian, but it is not possible to calculate PRR for all the DECs and is less specific.
Reporting Odds Ratio (ROR)	$\frac{M/P}{N/Q}$ $95\%CI = e^{\ln(ROR)\pm 1.96\sqrt{\frac{1}{M}+\frac{1}{N}+\frac{1}{P}+\frac{1}{Q}}}$	95% CI > 1 n ≥ 2	ROR provides the reporting odds of the suspected ADR of interest among all the ADRs associated with the suspected drug, comparing to the reporting odds for all the drugs included in the database. Application and interpretation of ROR is easier and the method is more sensitive than Bayesian as well as different adjustments for covariates in logistic variation analysis, but it is less specific and calculation is not possible if zero is the denominator.
BAYESIAN METHODS			
Multi-item Gamma Poisson Shrinker (MGPS)	$\frac{M(M+N+P+Q)}{(M+P)(M+N)}$	$EBGM_{05} > 2$ n > 0	It can be validly used to any DECs and is highly specific when threshold criteria are met. The Bayesian statistics are often complicated to use and are less sensitive.
Bayesian Confidence Propagation Neural Network (BCPN)	$\log_2 \frac{M(M+N+P+Q)}{(M+P)(M+N)}$	IC - 2 SD > 0	It can be applicable to any DECs and is highly specific when threshold criteria are met. The Bayesian statistics are often complicated to use and are less sensitive. In higher dimensions, it can be used forpattern recognition.

4. Application of Artificial Intelligence

Artificial Intelligence (AI) / machine learning provides a scope to improve the existing quantitative methods of signal detection by overcoming the disadvantages of these methods and obtaining true signals. AI can be used to detect the signals as early as possible and implement risk minimization measures promptly to ensure optimal patient safety. Data mining is time consuming process and the use of AI can reduce the analyzing time and human efforts in the signal assessment. AI is essential in the adjustment of the disproportionality score by removing noise such as confounding factors, missing data etc. and calculating observed/expected ratio to generate true and meaningful signals[7].

5. Conclusion

Disproportionality analysis mainly helps us in identifying the statistical associations involving suspected pharmaceutical products and adverse events by using the respective databases of safety reports obtained by spontaneous reporting. These spontaneous data consist of confounding factors which can interfere with the generation of the signal. Hence the clinically relevant data can increase the strength of clinical data by checking the factors such as time to onset, medical history, disease morbidity and de-challenge or re-challenge. Large amount of spontaneous data require further development of quantitative methods along with Artificial Intelligence (AI) in order to identify the true positive signals and avoid the false positive signals.

REFERENCES

1. Bate, A., Lindquist, M., Edwards, I., Olsson, S., Orre, R., Lansner, A. and De Freitas, R. (1998). A Bayesian neural network method for adverse drug reaction signal generation. European Journal of Clinical Pharmacology, [online] 54(4), pp.315-321. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9696956.
2. Bousquet, C., Henegar, C., Louët, A., Degoulet, P. and Jaulent, M. (2005). Implementation of automated signal generation in pharmacovigilance using a knowledge-based approach. *International Journal of Medical Informatics*, [online] 74(7-8), pp.563-571. Available at: https://www.ncbi.nlm.nih.gov/pubmed/15955732.
3. Dumouchel, W. (1999). Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA
4. Spontaneous Reporting System. *The American Statistician*, [online] 53(3), pp.177-190. Available at: https://www.tandfonline.com/action/showCitFormats?doi=10.1080%2F00031305.1999.10474456.
5. Elisabetta, P., Emanuel, R., Carlo, P. and Fabrizio, D.P. (2012). Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS). *Data Mining Applications in Engineering and Medicine, Adem Karahoca, IntechOpen*.DOI: 10.5772/50095. Available from: https://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine/data-mining-techniques-in-pharmacovigilance-analysis-of-the-publicly-accessible-fda-adverse-event-reporting-system-(AERS)/
6. Evans, S., Waller, P. and Davis, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 10(6), pp.483-486.
7. *In Vivo*. (2020). *The Role Of AI In Signal Detection*. [online] Available at: https://invivo.pharmaintelligence.informa.com/IV124409/The-Role-Of-AI-In-Signal-Detection.
8. Klservblog.com. (2020). *Pharmacovigilance | Drug Safety | Safety Assessments | Regulatory Affairs | Medical Information*. [online] Available at: https://klservblog.com/2018/12/03/artificial-intelligence-for-quantitative-signal-detection-methods-in-pharmacovigilance-and-drug-safety/
9. Neha, R., Subeesh, V., Beulah, E., Gouri, N. and Maheswari, E., 2019. Existence of Notoriety Bias in FDA Adverse Event Reporting System Database and Its Impact on Signal Strength. Hospital Pharmacy, p.0018578719882323.
10. Subeesh, V., Maheswari, E., Singh, H., Beulah, T.E. and Swaroop, A.M., 2019. Novel adverse events of iloperidone: a disproportionality analysis in US Food and

Drug Administration Adverse Event Reporting System (FAERS) database. Current drug safety, 14(1), pp.21-26.

11. U.S. Food and Drug Administration. (2020). *Data Mining at FDA -- White Paper*. [online] Available at: https://www.fda.gov/science-research/data-mining/data-mining-fda-white-paper
12. Van Puijenbroek, E., Bate, A., Leufkens, H., Lindquist, M., Orre, R. and Egberts, A. (2002). A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiology and Drug Safety*, [online] 11(1), pp.3-10. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11998548.
13. Who-umc.org. (2020). *UMC What is a signal?* [online] Available at: https://www.who-umc.org/research-scientific-development/signal-detection/what-is-a-signal/
14. World Health Organization. (2020). *Pharmacovigilance*. [online] Available at: https: // www. who. int /medicines / areas / quality_ safety /safety_ efficacy / pharmvigi / en /
15. Xiao, C., Li, Y., Baytas, I., Zhou, J. and Wang, F. (2018). An MCEM Framework for DrugSafety Signal Detection andCombination from Heterogeneous Real World Evidence.*Scientific Reports*, 8(1)