

BIG DATA ANALYTICS IN PHARMACOVIGILANCE - A GLOBAL TREND**SWETHA RANI AITHA¹, SRAVANI MARPAKA^{2*}, CHAKRADHAR T¹, BHUVANESHWARI E³,
SWARUPA RANI KASUKURTHI¹**¹Department of Pharmacology, Osmania Medical College, Koti, Hyderabad, Telangana, India. ²Adverse Drug Reaction Monitoring Centre-Osmania Medical College, Koti, Pharmacovigilance Program of India, Indian Pharmacopoeia Commission, Hyderabad, India.³Department of Pharmacology, Gandhi Medical College, Musheerabad, Secunderabad, Telangana, India. Email: sravani.marpak21@gmail.com*Received: 08 July 2021, Revised and Accepted: 24 August 2021***ABSTRACT**

Big data analysis has enhanced its demand nowadays in various sectors of health-care including pharmacovigilance. The exact definition of big data is not known to many people though it is routinely used by them. Big data refer to immense and voluminous computerized medical information which are obtained from electronic health records, administrative data, registries related to disease, drug monitoring, etc. This data are usually collected from doctors and pharmacists in a health-care facility. Analysis of big data in pharmacovigilance is useful for early raising of safety alerts, line listing them for signal detection of drugs and vaccines, and also for their validation. The present paper is intended to discuss big data analytics in pharmacovigilance focusing on global prospect and domestic country-India.

Keywords: Big data, Computerized medical information, Health-care sectors, Analysis, Pharmacovigilance.

© 2021 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2021v14i10.42765>. Journal homepage: <https://innovareacademics.in/journals/index.php/ajpcr>

INTRODUCTION

Big data is marked and cited as keyword nowadays in various sectors of health care including pharmacovigilance (Pv). The exact definition of big data is not known to many people though it is routinely used by them. An illustrious definition of big data relates to enormous data which are generated and are readily available for use. Increasing data availability along with technological advances provide new possibilities to store, mine, and analyze data across multiple data sources.

Big data refer to immense and voluminous computerized medical information which are obtained from electronic health records, administrative data, registries related to disease and drug monitoring [1]. This data is usually collected from doctors and pharmacists in a health care facility. However, the main drawback of this data is that its correctness has to be checked as there may be some inter-related issues such as duplications and so on [1]. In ancient times, data was collected, written, and stored in the form of records (Table 1). As time passed, these records were damaged and the information about the healthcare events was not available. With the advent of big data in health-care facility every known information is stored in a database and is made available for use. Despite its availability, for a long time, the true use of big data for analysis is popularized in recent times. The use of big data analytics was started increasing in Europe, America, and recently in Asia.

Pv involving adverse drug events (ADEs) and adverse drug reactions (ADRs) became the knowledge of interest with the evolution of new drugs and vaccines by various pharmaceutical companies. Moreover, manual recording of all these events and reactions was a difficult and troublesome task. The introduction of big data analysis in Pv bridged all the gaps to the regulatory authorities, pharmaceutical companies, and researchers by providing all relevant information related to data collection, assessment of ADE/reaction, signal detection, and validation.

This article gives an overview concept of present big data mining methods, approaches, applications, and limitations along with ethical considerations and future prospects in Pv.

SOURCE OF INFORMATION AND SEARCH STRATEGY FOR IDENTIFYING RELEVANT LITERATURE:

To ensure a comprehensive research review of the study, search terms were carried out using key text words as "Big Data," "Data Analytics," "Data Mining," "Real and Smart Data," "Drug safety," and "Pv." These terms were used individually and in combination to sew up for widening the search window of the literature collection. Relevant published articles from data bases such as "PubMed, Medline, and SCOPUS," were selected for disusing the objective of the review.

BIG DATA IN PV

Big data is used in Pv to analyze the large electronic information about ADEs/reactions which may be structured or unstructured in spontaneous reporting system databases and other databases [2]. Post-marketing surveillance of drug safety is mainly reviewed by systematic analysis of spontaneous reporting system database [3] (Table 2). Vigiflow-UMC, ADRMS (in developing stage), Safe-Vac for AEFIs in India, Eudra-vigilance in the European Union and FAERS (FDA Adverse Event Reporting System) in the United States, and JADER (Japanese ADE Report) in Japan are few spontaneous reporting system databases in Pv [4]. Vigibase is the global Pv monitoring database maintained and monitored by the World Health Organization-Uppsala Monitoring Centre located in Sweden [5].

HOPE, HYPE, AND REALITY OF BIG DATA IN PV

The hope and much of the hype of big data for Pv center on new data streams and technologies as a source for identifying potential new safety signals [6]. Electronic health records are also one of the big data used in Pv. Other big data used in Pv include databases of hospital discharge, out-patient diagnostic tests, laboratory test findings, emergency department, pediatrics and geriatric assessment records, mortality, and disease registries [7]. Social media is also used to promote Pv. This media uses text mining and is in the stage of development [8]. Using social media such as Twitter or Facebook, several patients share their personal experiences related to drug therapy and provide a good source for early signal detection [9].

All this computerized electronic information is processed and assessed for signal detection of various ADRs [10] (Fig. 1). Signal detection is the process of actively searching and identifying safety information about ADRs from a wide variety of data sources [11]. Data mining is one of the common data analysis methods for signal detection. Other common methods include Geographic Information systems (GIS), Text and information mining (ADETM), and Visualization tools [12]. Among the various statistical techniques covered in data mining Pv analytics will use the following:

- Descriptive modeling: This method is used to uncover shared similarities or grouping in historical data to determine reasons behind success or failure. Techniques belonging to this group include clustering, anomaly outlier detection, association rule learning, principal component analysis, and affinity grouping [12]
- Predictive modeling: This method is used to classify events in the future or estimate unknown outcomes. Techniques belonging to this group include regression, artificial neural networks, decision trees, random forest, and support vector machines [12]
- Prescriptive modeling: This method is also called the last frontier of analytic capabilities. It takes information from descriptive and predictive analytics and combines it with information obtained from unstructured data for improved prediction accuracy [12]
- Disproportionality methods: This method is used to identify the statistical association between products and events. These methods compare the observed count for product-event combination with an expected count [12].

Many data mining software are commercially available for Pv such as Empirica Signal, PV-Analyzer, SAS (Statistical Analysis System), and Molecular Analysis of Side Effects [13]. In the United States, Patient-Centered Outcome Research Network data system is used. Similar networks for patient safety have been developed around the world: Asian Pharmacoepidemiology Network in Asia [14] and Canadian Network for Observational Drug Effect Studies in Canada [15]. Networks such as Observational Medical Outcomes Partnership and Observational Health Data Sciences and Informatics, primarily focus on method testing and informatics tool development for data networks [16]. The joint International Society of Pharmacoepidemiology-International Society of Pharmacoeconomics and Outcomes Research taskforce on Real-World Evidence in health-care decision-making guides on the design and reporting of Pharmacoepidemiologic analyzes of longitudinal health-care databases [17-19].

All the above methods of big data analysis are used in various domains of Pv such as quality management, risk management, management of ICSRs, aggregate reporting, and signal management. These methods have demonstrated their promise in enhancing multiple areas of disease management, population health, and precision medicine. These augment clinical-decision making in healthcare. In the coming years, ongoing initiatives, such as the IMI's [26] WEB-RADR, and the Exponential Moving Average's goal to measure the impact of Pv practices, will identify the best uses of these data methods for Pv, along with patient populations, outcomes, or medicines that are best suited for signal detection. The outcome of any contribution of big data to Pv practice critically evaluates the impact of innovative data sources, techniques and whether these should uniquely complement or replace existing approaches, or are redundant, adding little, or no value to current Pv practice.

ROLE OF REAL-WORLD DATA IN THE ANALYSIS OF BIG DATA IN PV

Real-world data can include any information that is generated from the patient's records while admitted to the hospital and is treated to be the first resource when building predictive models. This data can also be generated from registries and pharmaceutical trials that the patient is undergoing for treatment or clinical trial studies. A well-collected repository of real-world data aids in accumulating enough weight for real-world evidence. Real-world evidence is to be considered for validating the big data models that are created to confirm their usefulness in Pv [27].

APPROACH TO BIG DATA IN PV

The big data approach to Pv [28] is performed in a systematic manner involving the following steps:

- Characterize relevant sources of big data and define the main format in which they can be expected to exist in
- Identify areas of usability and applicability of data
- Describe the current status of expertise, future needs, and challenges by gap analysis
- Generate a big data roadmap and list of recommendations.

The above steps will enhance the best way to analyze big data with accuracy by reducing manual errors and the time required for analysis [28]. Fig 2 depicts that relevant sources of big data may be derived from the web and social media (Facebook, Twitter, LinkedIn, Google, and WhatsApp), health-care databases in domestic country India (SOCH for HIV, Nikshay & Aarogyasethi for TB, and Cowin for COVID-19) are linked to drug safety information (PvPI-Pv Program of India, CDSCO-Central Drug Standard Control Organization, WHO-World Health Organization, and Safe-vac for adverse events following immunization [AEFI]). Processing and analysis are done using various IT tools [12] in big data such as Hadoop, R, STATA [27], Hbase, and Cassandra. Hadoop tool is mainly used to store, process, and analyze big data sets. Cassandra tool provides high availability and scalability. STATA explores the relationship between unstructured text to structured data using quantitative and qualitative methods [27]. After data collection and linkage from the above mentioned sources followed by processing and analysis Pv safety measures are implemented through education on safety use of medicines to patients, caution on prevention of ADRs/ADE from the regulatory agency is provided to health-care professional (HCPs) and patients, early detection of ADRs/ADEs and prediction of ADR occurrence. The advantage of linking data collection databases and records as mentioned above will help to analyze the proposed metrics applicable to all AEs detection, prediction and highlighting the designated medical events, and targeted medical events. These ICSR metrics are pooled using the MedDRA for medical coding of adverse events, WHO-DD for drugs and causality scale for causality assessment and expectedness from labels and literature, severity from severity scoring scales, and outcome of related reported events [27]. All these summarizing together with big data pooling would act as indicators for HCPs in implementation of early detection and prevention of AE(s) as triggering tools for patient safety in public health system either in domestic country (India) or globally. Big data offer BDD concept of "Breath, Depth, and Diversity." Breath includes large number of individuals making closer to the underlying source population with reduction in selection bias, depth relates to the increasing amount of data on each individual increases the chance that might have measures of likely confounders with potential reduction in information bias, and diversity relates to the various types of data that offer the potential to cross refer finding for any particular data source from datasets with significance to enhance control for residual bias [27,28].

USFDA also proposed a new surveillance program within its office of "Clinical Pharmacology (OCP)" which is known as "Pharmacological Mechanism-Based Drug Safety Prediction (PMDSP) [29]" program which uses the big data analysis to analyze the data sets of AEs reported and raises the safety alerts and communicates to the stakeholders to line list them for monitoring, reporting, and signal implementation for patient safety with the help of predictive tools [29]. PMDSP along with center for drug evaluation and research used the real-world evidence

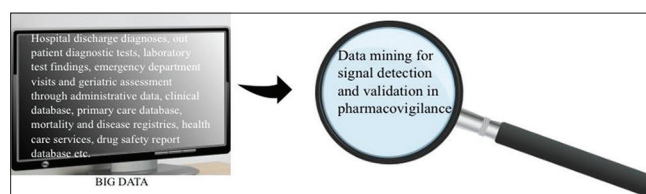


Fig. 1: Data mining using big data in



Fig. 2: Systematic approach to big data in pharmacovigilance

Table 1: Various safety reporting system databases

VIGIFLOW, ADRMS, Safe-Vac	INDIA
EUDRA-VIGILANCE	EUROPEAN UNION
FAERS	UNITED STATES
JADER	JAPAN
DEAN	AUSTRALIA

data to monitor and evaluate the safety of drug products after they are approved (post-market). Examples of such safety evaluation of real world data using large data sets of USFDA are risk of stroke after using antipsychotics, risk of seizures after using ranolazine, and risk of venous thromboembolism after an extended or continuous cycle of oral contraceptives. Similarly, it has also launched other program in connection with the Center for Biologics Evaluation and Research (CBER) focusing surveillance efforts within the sentinel system for vaccine safety, using the Post-Licensure Market Rapid immunization safety Monitoring (PRISM) system, for blood components with blood surveillance continuous active surveillance network (BloodSCAN) and initiated “Biologics Effectiveness and Safety (BEST)” [30] project to assure the safety and effectiveness of biologic products including vaccines, blood and blood products, human tissues, cellular products,

gene therapies, allergenics, xenotransplantation products, and devices related to biologics. Specially, it is using the BEST project to evaluate the large volumes of data sets in connection with rates of adverse events of special interest (AESI) for COVID-19 vaccine safety surveillance and monitoring. The project according with “CBER-BEST” was to mine data and recognize patterns through the use of technology to find as yet unrecognized safety signals after emergency use approval of COVID-19 vaccines and implement the benefit-risk plan [30].

APPLICATIONS OF BIG DATA ANALYTICS IN PV

Big data analysis performs safety assessments, comparative effectiveness studies, and investigational trials in a real-world setting. The following are the various applications of big data analytics in Pv:

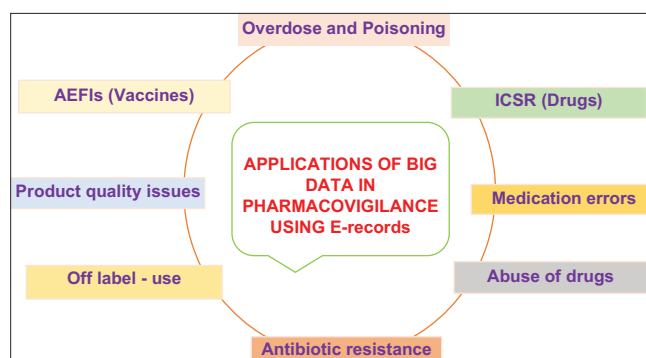
- Pre-marketing drug safety surveillance: Big data analysis is used to analyze reports, improve the identification process, and layout potential safety issues. Recent methods of big data analytics are also proved to be useful in research. Data mining saves time by the implementation of automated standardization of data, statistical scores, and prioritization of signals and also helps in understanding the biological basis for signals [31]
- Post-marketing safety surveillance: Pharmaceutical industries use big data analytics in identifying drug safety signals earlier, assessing risk, and interpreting clinical trial results [32]. Big data analytics save

Table 2: Recent methods of big data analytics in pharmacovigilance [20-25]

Artificial Intelligence	It consists of a wide range of technologies including learning, self-correction, and following rules [20]
Machine learning	Computer accesses data and use it to automatically learn and improve by experience [21]
Cognitive computing	It is useful in multiple domains of pharmacovigilance for the elimination of human error and standardization of processes [22]
Neural network	This is modeled according to the neuronal structure of the mammalian brain
Machine translation	In this method, the computer translates texts from one language to another
Natural language processing	This aids the computer to understand human language and language-related tasks
Semantic search	This method improves accuracy by understanding the searcher's intent and contextual meaning of terms
Block chain	Blocks are a list of records that are linked and secured using cryptography
Data mining	It is the process of extracting and transforming relevant information from unstructured data which can be used for further analysis and process
Natural language generation	This process transforms data into a written or narrative form and the output is generated from structured data
Autonomous software	It is the software that operates on behalf of users by employing some knowledge
Robotic process automation	This process utilizes software to perform traditional manual activities containing high volume and repetitive processes involving structured data
Desktop automation	It is the automation process within the desktop to provide guidance or assistance to human resources on demand
Predictive analytics and predictive reasoning	It is an advanced analytical process that uses current and historical data to draw inferences
Bots	These are the programs that carry out robotic process automation
Chatbots	These are bots designed to simulate human conversation through audio and text methods
Advanced analysis	It is an automated or semi-automated analysis process using sophisticated tools such as machine learning and neural network.
Image recognition	The computer system identifies objects, places, people, and writing in images using cameras, machine vision, and artificial intelligence
Machine vision/ computer vision	This method enables the computer to recognize objects for decision-making and additional processing

a great deal of time, money, and logistically helps the pharmaceutical industry in assessing signals in large volumes of data sets

- Regulatory decision making: Regulatory agencies rely on big data analytics for the identification of signals in cases of ADR and also in vaccine vigilance. Once a signal is detected using big data analysis, regulatory agencies make decisions concerning the implementation of signals as label changes, and benefit-risk monitoring of drugs [33]
- Clinical practice: In clinical practice, this analysis provides information on the disease, previous consultations, diagnostics, test results, and treatment. The blood type, allergies, diseases, possible medications, and vital signs measurement, everything are centralized and can be searched at a glance. Using big data analytics

**Fig. 3: Big data mining in pharmacovigilance-hospital setup**

in daily practice, correct information can be provided quickly in case of emergency

- Clinical trials: Advanced analytics can create value in clinical trials in the following ways:
 - Improved data quality by analyzing different kinds of data through automation
 - Efficient data mining by identifying deviations, differences, and improving transparency
 - Data-driven decision-making through corrective and preventive measures of data exploration.

Achievements of big data in Pv

- Danish and Swedish medical birth registries were used to study the prevalence of congenital malformations among infants exposed and not exposed to varenicline *in utero* by linkage to nationwide registries of dispensed prescriptions and hospital admissions [34]
- In the United States, the safety of meningitis B vaccination that is Trumenba vaccine exposure during pregnancy was studied using electronic health-care data and linked birth certificates from multiple health-care systems [35]
- In the United Kingdom, the researchers used smartphones and linked mobile data to study the relationship between weather patterns and rheumatoid arthritis symptoms. This study collects data about the severity of pain symptoms from an app and then links it to weather information based on the patient's location at the time of data entry [36]
- Pharmacoepidemiology research in the use of consumer wearable technology is gaining importance nowadays. These data streams and networks of sensors focus on medical and behavioral data such as heart rate, smoking status, alcohol consumption, and exercise pattern to examine relationships between data types previously unknown [37].

RECENT APPLICATIONS OF BIG DATA ANALYTICS IN PV

- **COVID-19 drugs and vaccines**
In COVID-19 pandemic, Pv analysis using big data plays a key role in reducing false information and providing correct data and explanations to patients about drug usage [38]. This analysis generates valuable information about a disease such as hidden patterns, unknown trends, correlations, and patient preferences. Key applications of Pv analytics in COVID-19 include the following:
 - Enhanced drug and vaccine safety and faster compliance
 - Benefit-risk analysis
 - Better protection for patients.

Big data analysis using data mining tools help to detect the new and unknown AEFI and AESI in subpopulation that is not adequately studied or excluded from trails, so continuity to monitor the safety concerns of these vaccines once they are rolled out with emergency use approval in connection to COVID-19 is on huge demand. Automated signals detected masking the noise of the background raised with the help of big data sets could be beneficial in analyzing the risk.

- Orphan drugs

In the past treatment of rare diseases was a challenge and there was little interest to develop new medicines for these diseases due to fewer market incentives. Since the number of people using orphan drugs is very small, it is difficult to conduct Pv. To solve this issue, patient support programs were established that help to produce safety reports too [39]. Analysis of these reports is of great importance in the Pv of orphan drugs.

DATA PRIVACY OF BIG DATA ANALYTICS IN PV

Data privacy is an important issue whenever the information is posted and allowing easy access for use of this data in research. The user is responsible for preserving the anonymity of information about the patient's identity. There are still ethical issues in using social media for the evaluation of new medical knowledge, the most important being each patient to be guaranteed to keep his anonymity. Current principles and guidelines for protecting individual's privacy rights lack technological developments. Procedures were designed to guarantee that the only data going through a step of data minimization should be accessible for analysis by registered end users, but the raw data were kept accessible in very specific circumstances to allow contacting the patient if drug withdrawal was required for safety reasons. Concerning privacy and confidentiality, much work is needed in terms of formulating guidelines to help drug developers understand the nature and extent of big data that would be deemed admissible in the drug approval process [40,41].

LIMITATIONS OF BIG DATA ANALYTICS IN PV

Despite big data play a major role in various sectors of Pv, there are certain limitations. These limitations include paucity of established standards and validation methods, bias, false signals, confounding variables, and data inconsistencies [42]. Available data which are incorrect, missing, or duplicate are another major challenge faced in big data analytics. Poor quality data algorithms may not be well understood by decision-makers to make inferences about particular signal detection. In India, the major challenge is faced mainly when patient's care may be documented in more than one electronic health record if they seek care at different institutions or practices leading to large amount of unstructured data and structured data might not substantially augment administrative data. Prescription versus dispensing challenges are also observed due to varied patient preferences.

LEGAL AND ETHICAL CONSIDERATIONS OF BIG DATA ANALYTICS IN PV

Legal considerations

Legal aspects in big data analysis are considered in terms of reporting. Reporting in Pv is performed through adverse event case reports, health agency databases (Vigibase, and FAERS), data from cohort studies, published literature, social media, and so on. These reports allow regulatory agencies to inquire into health records, electronic medical records, registries, and safety issues.

Ethical considerations [43]: Ethical aspects to be ensured while dealing with big data analysis in Pv include the following;

- Generating qualitative data of ADRs and their related treatment
- Transparency in maintaining and reporting ADRs
- Anticipate decision-making in the case of signal detection
- Establishment of risk-benefit profile based on the analysis of available data
- Identification of high-risk vulnerable populations.

FUTURE PERSPECTIVE OF BIG DATA ANALYTICS IN PV

Big data can identify correlations of patients' information within diverse datasets. Future directions which will overcome the current limitations of big data analytics in Pv include improvement inconsistency among data sources, validation of utilizing data sources, the establishment of standards for signal detection, application of an integrative approach

to signal detection, improvement of data mining software and tools, application of data mining to other product safety and regulatory issues [44].

CONCLUSION

Big data analytics has been successful in identifying new drug-related ADE/ADR for drug safety surveillance purposes. Although numerous challenges remain, the opportunity for big data to make further contributions to Pv efforts is evident. Improved methods, tools, and data sources used in drug safety surveillance are still in the early stages of development and are likely to further advance the use of big data in Pv. Finally, how well big data improve the detection of drug safety issues will be the true measure of its value.

ACKNOWLEDGMENTS

We acknowledge Osmania Medical College and NCC-PvPI, Indian Pharmacopoeia Commission, Ministry Of Health and Family Welfare, for technical support.

AUTHOR CONTRIBUTIONS

Dr. Swetha Rani Aitha contributed for drafting the manuscript, Ms. Sravani Marpaka has planned and designed the concept of manuscript, supported in drafting the manuscript and reviewed the manuscript, Dr. Chakradhar T and Dr. Bhuvaneshwari E reviewed the manuscript, Dr. Swarupa Rani Kasukurthi collected the reference articles.

CONFLICTS OF INTEREST

None declared.

FUNDING SOURCE

No source of funding.

REFERENCES

1. Harpaz R, Dumochel W, Shah NH. Big data and adverse drug reaction detection. *Clin Pharmacol Ther* 2016; 99:268-70.
2. Walker AM, Zhou X, Ananthakrishnan AN, Weiss LS, Shen R, Sobel RE, *et al*. Computer-assisted expert case definition in electronic health records. *Int J Med Inform* 2016; 86:62-70.
3. Trifiro G, Patadia V, Sturkenboom M. Post-marketing safety surveillance: Where does signal detection using electronic healthcare records fit into the big picture? *Drug Saf* 2013; 36:183-97.
4. Patadia VK, Coloma P, Schuemie MJ, Herings R, Gini R, Mazzaglia G, *et al*, EU-ADR Consortium. Using real-world healthcare data for pharmacovigilance signal detection-the experience of the EU-ADR project. *Expert Rev Clin Pharmacol* 2015; 8:95-102.
5. Uppsala Monitoring Centre. What is Vigibase? Available from: <http://www.who-umc.org/vigibase/vigibase>. [Last accessed on 2021 Jun 06].
6. Bate A, Reynolds RF, Caubel P. The hope, hype, and reality of big data for pharmacovigilance. *Ther Adv Drug Saf* 2018; 9:5-11.
7. Ventola CL. Big data and Pharmacovigilance: Data mining for adverse drug events and interactions. *PT* 2018; 43:340-51.
8. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection through multi-corpus training. *J Biomed Inform* 2015; 53:196-207.
9. De Rosa M, Fenza G, Gallo A, Gallo M, Loia V. Pharmacovigilance in the era of social media: Discovering adverse drug events cross-relating Twitter and PubMed. *Future Gen Comput Syst* 2021; 114:394-402.
10. Bate A, Pariente A, Hauben M, Begaud B. Quantitative signal detection and analysis in pharmacovigilance. *Manns Pharm* 2014;26:159-86.
11. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc* 2014; 21:353-62.
12. Jose Rossello. Available from: <https://www.cuelogic.com/blog/big-data-approaches-in-pharmacovigilance>. [Last accessed on 2021 Jun 13].
13. Duggirala HJ, Tønning JM, Smith E, Bright RA, Baker JD, Ball R, *et al*. Use of data mining at the food and drug administration. *J Am Med Inform Assoc* 2016; 23:428-34.
14. AsPEN collaborators; Andersen M, Bergman U, Choi NK, Gerhard T,

- Huang C, Jalbert J, *et al.* The Asian pharmacoepidemiology network (AsPEN): Promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; 22:700-4.
15. Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, *et al.* CNODES: The Canadian network for observational drug effect studies. *Open Med* 2012; 6:e134-40.
 16. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, *et al.* Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers. *Stud Health Technol Inform* 2015; 216:574-8.
 17. Wang SV, Schneeweiss S. On behalf of the joint ISPE-ISPOR special task force on real-world evidence in health care decision making: Reporting to improve reproducibility and facilitate validity assessment for health care database studies V1.0. *Pharmacoepidemiology Drug Saf* 2017; 26:1570.
 18. Berger ML, Sox H, Willke R. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the Joint ISPOR-ISPE special task force on real-world evidence in health care decisionmaking. *Pharmacoepidemiol Drug Saf* 2017; 26:1033-9.
 19. Bate A. Guidance to reinforce the credibility of health care database studies and ensure their appropriate impact. *Pharmacoepidemiol Drug Saf* 2017; 26:1013-7.
 20. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med* 2019; 25:44-56.
 21. Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device. United States: Food and Drug Administration; 2019. Available from: <https://www.fda.gov/media/122535/download>. [Last accessed on 2021 Jun 01].
 22. Chen Y, Argentinis E, Weber G. IBM Watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther* 2016; 38:688-701.
 23. Simon Fraser University. Big Data Glossary. Available from: <https://www.tinyurl.com/y6vhvgff>. [Last accessed on 2021 May 26].
 24. Global Health Network. Global Pharmacovigilance. Glossary of Drug Safety Terms. Available from: <https://www.globalpharmacovigilance.tghn.org/resources/glossary>. [Last accessed on 2021 Jun 05].
 25. Big Data Made Simple. Big Data A to Z: A Glossary of Big Data Terminology; 2021. Available from: <http://www.bigdata-madesimple.com/big-data-a-to-zz-a-glossary-of-big-data-terminology>. [Last accessed on 2021 Jun 16].
 26. Klungel OH, Kurz X, de Groot MC, Schlienger RG, Tcherny-Lessenot S, Grimaldi L, *et al.* Multi-centre, multi-database studies with common protocols: Lessons learned from the IMI PROTECT project. *Pharmacoepidemiol Drug Saf* 2016; 25:156-65.
 27. Big Data Approaches in Pharmacovigilance: Using Big Real-world Data to Create Decision-relevant Evidence by Amith Govil; 2019. Available from: <https://www.cuelogic.com/blog/big-data-approaches-in-pharmacovigilance>. [Last accessed on 2021 Jun 06].
 28. Heads of Medicines Agencies (HMA)/European Medicines Agency (EMA) Joint Task Force Meeting on Big Data: Identifying Solutions for Big Data Challenges; 2018. Available from: <https://www.ema.europa.eu/en/events/heads-medicines-agencies-hma-european-medicines-agency-ema-joint-big-data-task-force-meeting>. [Last accessed on 2021 Jun 03].
 29. Abernethy DR, Woodcock J, Lesko LJ. Pharmacological mechanism-based drug safety assessment and prediction. *Clin Pharmacol Ther* 2011;89:793-7.
 30. USFDA-CBER Biologicals Effectiveness and Safety (BEST) System; 2020. Available from: <https://www.fda.gov/vaccines-blood-biologics/safety-availability-biologics/cber-biologics-effectiveness-and-safety-best-system>. [Last accessed on 2021 Jun 03].
 31. Yang CC, Jiang L, Yang H, Tang X. Detecting Signals of Adverse Drug Reactions from Health Consumer Contributed Content in Social Media. Beijing, China: Presentation at 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2012. p. 33-40.
 32. Arlett P, Kurz X, Fitt H. Increasing scientific standards, independence and transparency in post-authorization studies: The role of the European network of centres for pharmacoepidemiology and pharmacovigilance. *Pharmacoepidemiol Drug Saf* 2012; 21:690-6.
 33. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther* 2014; 96:239-46.
 34. Olsen M, Petronis KR, Froslev T, Mo J, Stephansson O, Granath F, *et al.* Maternal use of varenicline and risk of congenital malformations. *Pharmacoepidemiol Drug Saf* 2015; 24:244.
 35. Purcell B, Marie R, Griffin, Haynes K, Lin ND, McMahill-Walraven CN, *et al.* Safety of Trumenba vaccine among pregnant women in the United States: Planning and design of a large-scale multi-site observational study. *Pharmacoepidemiol Drug Saf* 2017; 26:410.
 36. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, *et al.* Cloudy with a chance of pain: Engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR Mhealth Uhealth* 2017; 5:e37.
 37. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: Promises and barriers. *PLoS Med* 2016; 13:e1001953.
 38. How is Pharmacovigilance Analytics Helping the Fight Against COVID-19, Shanawaz Sheriff; 2020. Available from: <https://www.acuvate.com/blog/pharmacovigilance-analytics>. [Last accessed on 10 Jun 2021].
 39. Joppi R, Bertele V, Garattini S. Orphan drugs, orphan diseases. The first decade of orphan drug legislation in the EU. *Eur J Clin Pharmacol* 2013; 69:1009-24.
 40. Brill J. Federal Trade Commission. Protecting Consumer Privacy in an Era of Rapid Change: A Discussion of the FTC's Privacy Report. Available from: <https://www.ftc.gov/public-statements/2012>. [Last available on 2021 Jun 10].
 41. European Parliament. Council of the European Union E. U. Directive, 95/46/EC of the European Parliament and of the Council of 24 October, 1995 on the Protection of Individuals Concerning the Processing of Personal Data and on the Free Movement of Such Data. Vol. 23. Brussels, Belgium: European Parliament; 1995. p. 348.
 42. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M, *et al.* Social media and pharmacovigilance: A review of the opportunities and challenges. *Br J Clin Pharmacol* 2015; 80:910-20.
 43. McKee R. Ethical issues in using social media for health and health care research. *Health Policy* 2013; 110:6.
 44. Korda M. 21st Century Cures Act: Back to the Future; 2021 Available from: <https://www.nbmecommunity.nature.com/users/21923-michelle-korda/posts/14629-21st-century-cures-act-back-to-the-future>. [Last accessed on 2021 Jun 10].