tcs ADD
Digital Advantage for Life Sciences

Building on belief

A Hybrid Approach towards Signal Detection and Signal Management for Pharmacovigilance

Abstract

The process of signal management in pharmacovigilance includes a set of activities, which aim to determine whether there are new risks associated with a particular medicinal product, or whether known risks associated with a particular drug have changed in frequency or severity. The signal management process follows a systematic approach: Signal detection, Signal Validation Signal prioritization, Signal assessment, Recommendation for action, and finally Exchange of information. Signal detection is the most vital part of signal management. Traditional methods include usage of solitary algorithms which are not accompanied by additional features of qualitative data mining resulting in mediocre performance. Intense efforts are invested to develop the quantitative and qualitative detection algorithms. To prevail over this limitation, we came up with a hybrid approach to effectively solve this drawback in which the algorithm is fueled to assess a signal, and also manage it for effective grading based on a qualitative approach. The present article provides a comprehensive and detailed description of our hybrid approach.

Introduction

Every pharmaceutical product approved to be used in a marked setting, has proven benefits, but also associated adverse effects. Timely detection of such unknown risks is pivotal to ensure the patient's safety. The detection process applies to all medicinal products, covering their entire life cycle, specifically including clinical development and post-market phases, for any type of adverse event, serious or non-serious. Signal detection and signal management in pharmacovigilance involves ongoing monitoring to identify case reports or case report series of adverse events (AE) that are worthy for further exploration and potentially requires safety actions such as a safety signal investigation.. Traditionally, signals are detected either qualitatively or quantitatively. The former involves the qualitative analysis through the manual assessment of Individual Case Safety Reports (ICSR) in an individual or cumulative manner.

The latter, on the other hand involves the more common quantitative approach that makes use of statistical techniques - the most often used being disproportionality analysis.

To identify the disproportionate reporting ratios the most common technique used is data mining. Data mining techniques most commonly known as Signal Disproportionality Analysis (SDAs) are used to explore a wide variety of databases of spontaneous reports for previously unknown associations between medicinal products and reported Adverse Events (AEs) that might have escaped the monitoring of manual case assessment. Quantitative signal detection or more specifically disproportionality analysis is most commonly done through disproportionality statistics i.e., by taking into account the ratio of the proportion of spontaneous ICSRs to the proportion that would be expected if no association had existed between the medicinal product and the event. A plethora of

different ways to calculate disproportionality are available, the most common or classical methods such as the proportional reporting ratio (PRR) or the reporting odds ratio (ROR) as well as using the Bayesian methods such as the Multi-item Gamma Poisson Shrinker (MGPS) and the Information Component (IC).

Most pharmacovigilance departments maintain a system to identify adverse drug reactions (ADRs) through analysis of spontaneous reports. The signal Disproportionality Analysis (SDAs) and the nature of the reporting databases vary between operators and it is unclear whether any algorithm can be expected to provide good performance in a wide range of environments. Although the current clinical assessment mostly relies on disproportionality analysis, it is solely based on aggregate numbers of reports and hence overlooks the quality and nature of content of the report.

Moreover, signal management process includes the following steps:

- Signal detection,
- Signal Validation
- Signal prioritization,
- Signal assessment,
- Recommendation for action, and
- Exchange of information

Typically, the traditional software's solutions lack a functional user interface for overall signal management. During the last 10 years it has been an exponential rise in data volume along with a proliferation of solicited and unsolicited safety reports, and predictions are that this trend will keep increasing over the next 5 years too.There will be vast increases and changes in surveillance data that will be reported in the near future. However, not only the volume of safety cases is increasing, also the source and type of records, including reports from electronic health records and claims, personal health records, standards for health data, data from Federal and private sector mobile devices for tracking health, and data from social websites (blogs, patient advocacy group sites, and search term logs). Hence a comprehensive Signal Management system would be required to handle such challenges.

Vigirank - a data-driven screening algorithm for identifying potential causally associated safety signals can be a good hybrid approach. It accounts for report quality and content along with disproportionate reporting. Signal detection and prioritisation is done by using this predictive Vigirank algorithm on the computed Vigirank variables. The computed variables are Disproportionality reporting ratio (specifically IC or Information Component), Recent Reporting, Geographic spread, Informative reports, Time to onset, Dechallenge, Rechallenge, Solely reported, and Multiple reporting elements. The algorithm has been implemented using Lasso Logistic Regression on the data made available from the FAERS dataset.

After identifying safety signals in an ad hoc manner, the next phase is signal prioritization. Vigirank provides an output score. This score along with a number of other factors can be taken for this phase. One can use reaction outcome or the seriousness features for prioritizing the signals with serious medical conditions. EMA has provided a set of serious events called Designated Medical Events (DMEs). This list acts as an important confounding factor for signal prioritization of signals. EMA suggests that the signals with these events should be assessed on high priority irrespective of the disproportionality ratios and, one should not entirely depend on this list alone. Important Medical Events (IMEs) is also a list of this kind which facilitates the prioritization task. Similarly every pharma company has their own list of events called Targeted Medical Events (TMEs) for their products.

Signal validation includes validating a signal both analytically and through existing literature. The literature in the form of research articles can be fetched in sites like PubMed, drug labelling information etc. Analytical charts give clarity on the data like gender distribution, percentage of reports with particular drug characteristics etc. This step is like an initial assessment for each detected signal. After validation, one may classify each signal into one of the 3 categories i.e., valid signal, not a signal and worthy of further analysis. According to EMA GVP module ix, the various considerations that are useful in this phase include

- previous awareness on the reaction and the available data on the SmPC of medicinal products
- strength of the evidence like disproportionality, quality of data, dose-reaction relationship etc.
- clinical relevance and context. Which include understanding of drug-drug reactions, severity and consideration of medication errors.

The solution presented thus provides various functionalities including but not limited to PubMed and Drug label literature review along with summarization functionality, annotations as well as graphs to illustrate various statistical figures. Sections like “Open signals”, “Closed Signals”, “Further Evaluation”, “Keep Under Monitoring” “Archive” are created to manage the flow of a signal and all these sections have tables representing all the signals present in that stage of signal management. Since the quality and content of individual reports is of fundamental importance, we make use of the Vigirank algorithm along with an efficient user interface thus aiming to integrate the value of automation with the breadth of aspects used in clinical assessment.

Datasets

While building the solution, numerous medical data was exhaustively analysed which was of paramount importance to prove the scalability, robustness, and superiority of the proposed system. Findings have been compiled below, subject to certain limitations.

a) FAERS is a public spontaneous adverse event reporting system by FDA. Its data is open.

b) EudraVigilance is also an open safety database from EMA.

c) Drug labelling information of each medicinal product is used in the signal validation phase

d) PubMed, Cochrane Library are used to verify existing research literature and reviews for a particular drug event combination.

Signal detection and management

A response to a drug which is noxious and intended, and which occurs at doses normally used in humans for the prophylaxis, diagnosis, or therapy of disease, or for the modification of physiological function. An active surveillance system is the collection of case safety information as a continuous pre-organized process. Spontaneous reporting is done by the system whereby case reports of adverse events are voluntarily submitted from health professional and pharmaceutical manufacturers to the national regulatory monitoring authorities.

Signal is a reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously. Adverse event can be classified into four parts

1) Severity : mild, moderate, or severe
2) Seriousness : Non serious or serious
3) Expectedness : Expected or Unexpected
4) Casuality : Related or Unrelated

Multiple factors contribute to Adverse Drug Reactions (ADRS)

1) Poor knowledge of pharmacology, adverse effects of drugs
2) Irrational use of drugs, poor prescribing patterns
3) Promotional activities by pharmaceutical company detailers
4) Lack of authentic sources of information
5) Liberal drug outlets and unhealthy pharmaceutical practices
6) Liberal OTC and self medications practices
7) Ignorant, illiterate public

Signal detection

Signal detection Reporter information on a possible causal relationship between adverse event and drug. The relationship between the adverse event and a drug is either unknown or incompletely documented previously. Usually more than a single report is required to generate a signal.. Depending on the attributes of a signal such as seriousness, reaction outcome of event and the quality of the information a signal is considered to be evaluated further. It cannot be regarded as definitive as an evaluation and validation of the signal is necessary . Traditional approaches use data mining algorithms to generate a signal.

Challenges and data-mining mitigations related to safety report databases are as follows:

- Missing, incorrect, or vague information
- Separate reports about the same incident
- Events may be due to the treated condition, another condition, or another product.
- Over-reporting
- Timeliness of reporting and processing

Signal management

Signal management processes is a set of activities performed to determine whether, based on an examination of individual case safety reports (ICSRs), aggregated data from active surveillance system or studies, literature information or other data sources there are new risks associated with an active substance or medicinal product or whether the known risk has changed.

Steps followed in signal management are:

1. Signal detection
2. Signal validation
3. Signal analysis and prioritisation
4. Signal assessment
5. Recommendation for action
6. Exchange of information

Traditional approaches

Although the present clinical assessment mostly focuses on disproportionality analysis, it does so purely on the basis of aggregate numbers of reports, ignoring the quality and character of the report's content. In typical circumstances, the lack of functional user interface for overall signal management in traditional software's solutions is a huge drawback. SDAs only consider quantitative parameters while ignoring the qualitative parameters which can have a role in determining if a given medicinal product and reaction combination is signal or not.

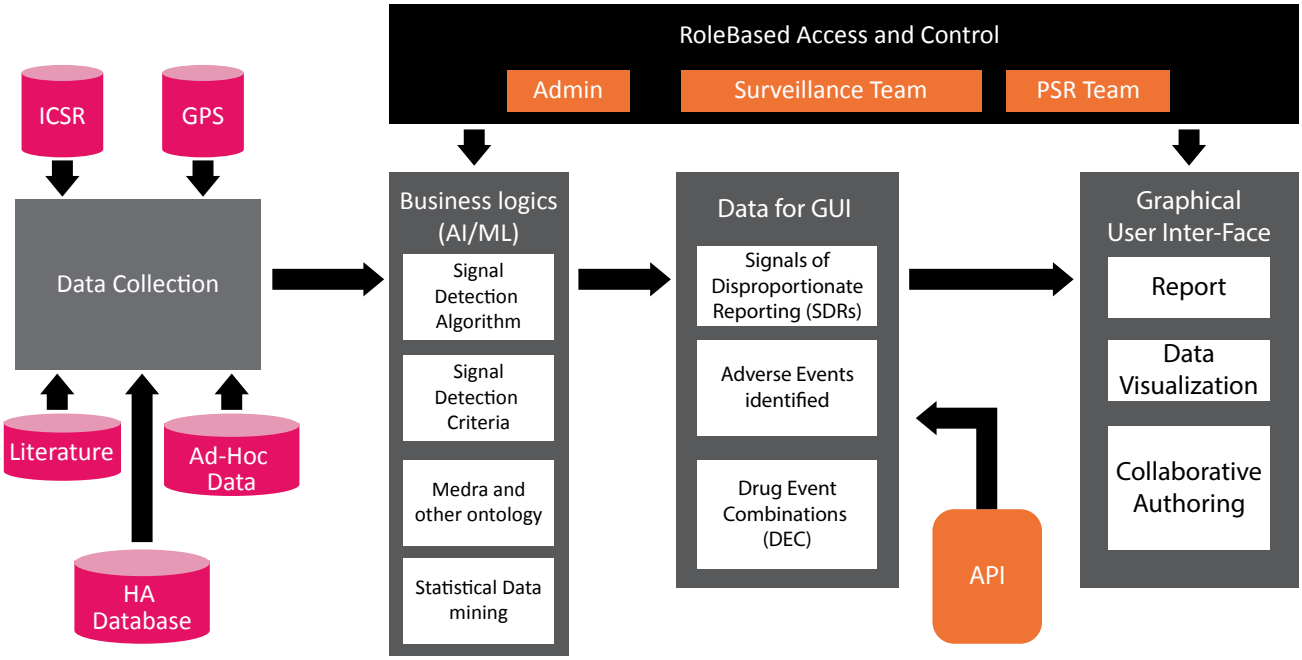

Fig. 1: Proposed System Architecture

Vigirank and vigigrade

It is a predictive algorithm based on the training, it predicts a score as output for each signal. Machine learning implemented using Lasso Logistic Regression. The required nine variables are computed from the database. All the computed variables are transformed using mathematical curves to maintain consistency and diminish the effect on addition of extra reports. The transformed variables are taken as inputs for the vigirank and a confidence score is generated as output. This score tells about the potential of a signal to be further assessed. The output score is used for signal prioritization. Vigirank Training: Known positive signals and negative signals are prepared as follows:

a) Positives: collected historical safety signals from project PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium) co-ordinated by EMA.

   (http://www.imiprotect.eu/adverseDrugReactions.shtml).

b) Negatives: are made by matching products with meddra PTs of HLTs which are not the HLTs of confirmed PTs for that product. For all the prepared signals, the nine variables are computed.

A target binary field is created representing whether the signal is positive or negative (i.e., 1 or 0 respectively). Available data is split into training and testing signals. Lasso Logistic Regression, a Machine learning algorithm is used to train the vigirank using training set.

Proposed system architecture

Public databases like FAERS, EudraVigilance etc have the safety reports in the form of ICSRs (i.eIndividual Case Safety Report) in E2B/M2 formats. The ICSRs are collected from the FAERS database. A total of 45 fields are collected from each ICSR and stored in the project database.

Known positive signals and negative signals are prepared as follows:

1. Positives: collected historical safety signals from project PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium) co-ordinated by EMA.

2. Negatives: are made by matching products with meddra PTs of HLT.

A target binary field is created representing whether the signal is positive or negative (i.e., 1 or 0 respectively). Available data is split into training and testing signals. Lasso Logistic Regression, a Machine learning algorithm is used to train the vigirank using training set.

The Vigirank algorithm is used for signal detection. Vigirank uses quantitative and qualitative variables for detecting and prioritizing the signals before in-depth assessment. Here a improved vigirank algorithm is applied using more independent variables like Disproportionality (IC), time to onset, dechallenge, rechallenge, Solely reported, multiple reporting system, informative reports, geographical spread, recent reporting.The nine variables are computed for a signal using the available data records. Here, Information component (IC) is a quantitative variable taken for the vigirank. This disproportionality algorithm is used by UMC (Uppsala Monitoring Centre). Any other algorithm can be used as a variable for vigirank in the same way. The results of Proportional Reporting Ratio (PRR), Reporting Odds Ratio (ROR), Gamma Poisson shrinkage (GPS) for each signal are considered with IC.

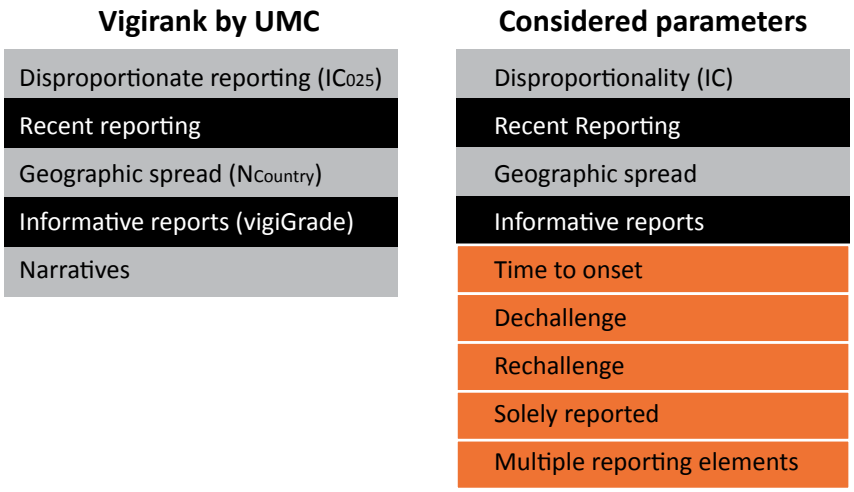

Fig. 2: Parameters considered by out proposed system

The transformed variables are taken as inputs for the vigirank and a confidence score is generated as output. This score tells about the potential of a signal to be further assessed.The output score is used for signal prioritization.

A priority score (ranging 1-12) and a priority remark is generated on the basis of DMEs, IMEs, vigirank output score, vigigrade, reaction outcomes and seriousness features.

DME : Designated Medical Events. The list includes serious medical events that are assessed on high priority irrespective of statistical importance.

IME : Important Medical Events. This list specifies important adverse events and helps in day to day pharmacovigilance activities by EMA

Signal validation includes analysis of different data fields for deeper insights. Here, charts like gender distribution, drug characterization, geographic spread through country map plot, forest plot for disproportionality analysis, node graphs etc are implemented.Relevant literature is fetched through an API to evaluate a signal. NLP functionalities such as text summarization and keyword extraction are integrated. Managing the flow of a signal from incoming signal to further analysis to archive/end.

“Open signals”, “Further Analysis”, “Archive” are created to manage the flow of a signal. ‘Open signals’ section contains all the signals that are yet to be assessed. After completing initial assessment, the user can write the necessary annotation, change the status and send to other sections. If the signal is found to be negative, they will be sent to the ‘Archive’ section. Otherwise, it will be sent to the ‘Further analysis’ section if the user wants a deeper assessment. Users can change the status of a signal anytime. Annotation of the signal by Master and Editor. Downloading detailed reports in the form of pdf and csv can be implemented. Dataset can also be downloaded. Access control is under Master’s supervision. Master can edit or remove a viewer.

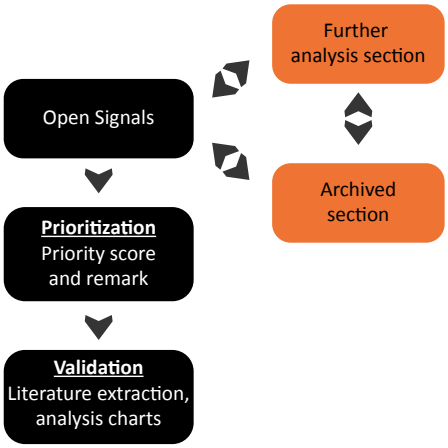

Fig. 3: Flow diagram for management of signals

Traditionally, signals are detected either qualitatively through the manual assessment of individual case safety reports (ICSR) or quantitatively through statistical techniques - the most commonly used being disproportionality analysis.Moreover, the traditional software solutions lack a functional user interface for overall signal management. The current method of detecting a signal is predominantly based on detection and validation with interactive graphical user interface for Signal Management. Our method is a hybrid method which assesses each signal with unambiguity and produces quality results.

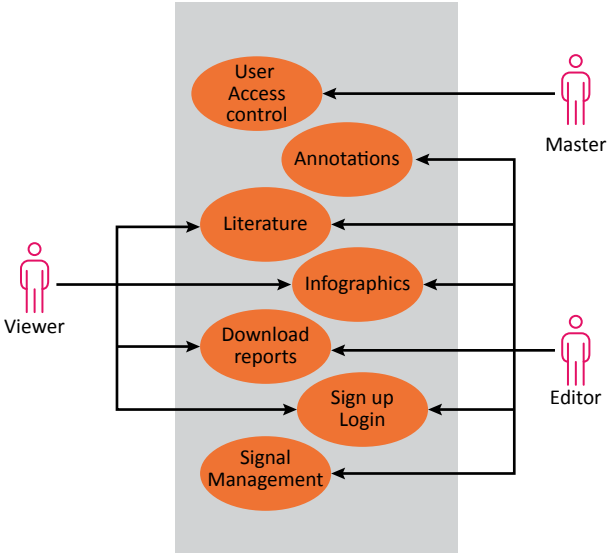

Fig. 4: Base Use case diagram

Future scope

Pharmacovigilance is an important and life saving field where Signal detection and its assessment has a paramount significance in ensuring that safe drugs are only served to the patient. Since clinical trials are conducted under strictly controlled conditions they only provide limited information to work on signal detection. Some of the adverse drug reactions can be detected only after long term use in larger populations and in specific patient groups due to specific concomitant medications or disease. Marketing the drug which has trust of people and known reactions are known in order to avoid any adverse event, the detection of safety signals which are not previously known needs to be explored as early as possible. For this post marketing data is used and it is one of the major challenges of pharmacovigilance. The current method of detecting a signal is predominantly based on detection and validation but the management part is still lagging. So here comes our hybrid method which assesses each signal with unambiguity and produces quality results.

1. The application can provide the option to run the vigirank algorithm with their desired variables. This can include their desired disproportionality algorithm.

2. NLP can be applied on the text obtained from literature extraction thus assisting the user in a better way.

3. Users can be tracked and any changes if present, shall be logged into the database.

4. Audit trail for a signal showcasing its entire journey with all the required attachments, dates can be implemented and submitted to an user in the form of a report.

5. Email and pop-up notifications can be enabled.

6. All the graphs and other analytics can be made dynamic

7. Drug-drug similarity can also be modeled using pharmacological data of several drugs. This will be helpful in signal evaluation.

8. A decision tree can be constructed to show the recommendations to the users about each signal.

9. Drug molecules and their pharmacokinetics/pharmacodynamics can be shown in graphical/ animated forms.

Conclusion

We came across signal detection, signal prioritization, signal validation, how signal is processed in native methods, its flaws and how we came through those difficulties using our hybrid methodology. Thus, the hybrid method's result precision is comparatively higher than the native method and hence we are one step closer to perfecting the signal processing. Though signal detection is our priority, signal management is also the major part we need to focus on. Our hybrid method incorporated major signal management techniques to produce unambiguous results to the user as well as coping up with the ever increasing sizes of the reports databases without compromising the speed of identification of potential safety issues and aid in prioritization safety issues. Free personnel can devote a higher proportion of their time to tasks that aren't yet readily assisted by machines.

References

[1] *Arnaud, M., Bégaud, B., Thurin, N., Moore, N., Pariente, A., & Salvo, F. (2017). Methods for safety signal detection in healthcare databases: a literature review. Expert Opinion on Drug Safety, 16(6), 721–732. https://doi.org/10.1080/14740338.2017.1325463*

[2] *Bate, A. (n.d.). Disproportionality statistics for signal detection. Retrieved May 23, 2021, from https://isoponline.org/wp-content/uploads/2015/09/Disproportionality-SD-Andrew-Bate.pdf*

[3] *Bergvall, T., Norén, G. N., & Lindquist, M. (2013). vigiGrade: A Tool to Identify Well-Documented Individual Case Reports and Highlight Systematic Data Quality Issues. Drug Safety, 37(1), 65–77. https://doi.org/10.1007/s40264-013-0131-x*

[4] *4) Bio.Entrez package — Biopython 1.75 documentation. (n.d.). Biopython.org. Retrieved May 23, 2021, from https://biopython.org/docs/1.75/api/Bio.Entrez*

[5] *Caster, O., Juhlin, K., Watson, S., & Norén, G. N. (2014). Improved Statistical Signal Detection in Pharmacovigilance by Combining Multiple Strength-of-Evidence Aspects in vigiRank. Drug Safety, 37(8), 617–628. https://doi.org/10.1007/s40264-014-0204-5*

[6] *Caster, O., Sandberg, L., Bergvall, T., Watson, S., & Norén, G. N. (2017). vigiRank for statistical signal detection in pharmacovigilance: First results from prospective real-world use. Pharmacoepidemiology and Drug Safety, 26(8), 1006–1010. https://doi.org/10.1002/pds.4247*

[7] *7) Guideline on good pharmacovigilance practices (GVP) Module IX -Signal management (Rev 1) Date of coming into effect of first version. (2017). https://www.ema.europa.eu/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-signal-management-rev-1_en.pdf*

[8] *Ismail, O. (2020, December 1). Scientific Article: The Worst Medical Disaster in Recorded History - TIU-Pharmacy Tishk International university. TIU-Pharmacy. https://pharmacy.tiu.edu.iq/scientific-article-the-worst-medical-disaster-in-recorded-history/*

[9] *List of DMEs. (2021). Europa.eu. https://www.ema.europa.eu/en/documents/other/designated-medical-event-dme-list_en.xlsx*

[10] *List of IMEs. (2018). Alims.gov.rs. https://www.alims.gov.rs/ciril/files/2018/10/Important-medical-event-terms-list-MedDRA-version-21_1.xls*

[11] *Magar, S. (2020, April 26). Signal Detection and Management. Pharmacovigilance Tutorials. https://pharmacovigilancetutorials.wordpress.com/2020/04/26/signal-detection-and-management/*

[12] *Patel, T. (n.d.). Pharmacovigilance and methods of Signal Detection. Retrieved May 23, 2021, from http://www.ijopp.org/sites/default/files/IJOPP_6_3_2013_16.pdf*

[13] *Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII • COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES. (n.d.). COUNCIL for INTERNATIONAL ORGANIZATIONS of MEDICAL SCIENCES. Retrieved May 23, 2021, from https://cioms.ch/publications/product/practical-aspects-of-signal-detection-in-pharmacovigilance-report-of-cioms-working-group-viii/*

[14] *PROTECT Home. (n.d.). Www.imi-Protect.eu. Retrieved May 23, 2021, from http://www.imi-protect.eu/adverseDrugReactions.shtml*

[15] *UMC | vigiMethods. (n.d.). Www.who-Umc.org. Retrieved May 23, 2021, from https://www.who-umc.org/vigibase/vigilyze/vigimethods/*

[16] *Vilar, S., Harpaz, R., Chase, H. S., Costanzi, S., Rabadan, R., & Friedman, C. (2011). Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. Journal of the American Medical Informatics Association, 18(Supplement_1), i73–i80. https://doi.org/10.1136/amiajnl-2011-000417*

[17] *Good Signal Detection Practices: Evidence from IMI PROTECT Antoni F. Z. Wisniewski1 • Andrew Bate2 • Cedric Bousquet3,4 • Andreas Brueckner5 • Gianmario Candore6 • Kristina Juhlin7 • Miguel A. Macia-Martinez8 • Katrin Manlik9 • Naashika Quarcoo10 • Suzie Seabroke11 • Jim Slattery6 • Harry Southworth12 • Bharat Thakrar13 • Phil Tregunno11 • Lionel Van Holle14 • Michael Kayser15 • G. Niklas Nore´n7*

[18] *https://www.fda.gov/science-research/data-mining/data-mining-fda-white-paper#intro*

About the authors

Shubham Chaudhari
Domain Consultant,
TCS ADD

Achyuth Tanay Kota
Domain Consultant,
TCS ADD

Ashutosh Pachisia
Data Scientist,
TCS ADD

Dr. Neelesh Joshi
Lead Life Sciences
NA

Dr. Alejandra Guerchicoff
Industry Leader –
TCS ADD Platform

Aakash Anand
Domain Consultant,
TCS ADD

Aditya Chouhan
Domain Consultant,
TCS ADD

Kamal Khumar
Domain Consultant,
TCS ADD

Surbhi Nair
Domain Consultant,
TCS ADD

Pratik Gada
Team Lead and
Safety surveillance
scientist Life
Sciences

Dr. Varun Vijay Vaidya
Lead Life Sciences UK
Project Management

TATA CONSULTANCY SERVICES

Awards and accolades

TOP 3
IT SERVICES
BRAND

FASTEST GROWING
IT SERVICES BRAND
FOR THE DECADE
2010 - 2020

About TCS ADD Platforms

TCS ADD is a modern and open drug development platform for life sciences that enables digital ecosystems, simplifies data complexity and provides faster access to new and effective drugs for patients in need. The platform is powered by our proprietary cognitive intelligence engine data driven smart analytics and Internet of Things (IoT) that makes clinical trials more agile and safe. TCS ADD leverages the best of cloud architecture and personalized user experience design in compliance with quality guidelines and privacy regulations.

To know more

Visit the https://www.tcs.com/tcs-add page on www.tcs.com

Email: add.platform@tcs.com

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is a purpose-led transformation partner to many of the world's largest businesses. For more than 50 years, it has been collaborating with clients and communities to build a greater future through innovation and collective knowledge. TCS offers an integrated portfolio of cognitive powered business, technology, and engineering services and solutions. The company's 500,000 consultants in 46 countries help empower individuals, enterprises, and societies to build on belief.

Visit **www.tcs.com** and follow TCS news **@TCS**.

Corporate Marketing | Design Services | M | 06 | 22