



Optimizing Signal Management in a Vaccine Adverse Event Reporting System: A Proof-of-Concept with COVID-19 Vaccines Using Signs, Symptoms, and Natural Language Processing

Guojun Dong¹ · Andrew Bate^{2,3} · François Haguinet² · Gabriel Westman⁴ · Luise Dürlich^{5,6} · Anders Hviid^{1,7} · Maurizio Sessa¹

Accepted: 14 November 2023
© The Author(s) 2023

Abstract

Introduction The Vaccine Adverse Event Reporting System (VAERS) has already been challenged by an extreme increase in the number of individual case safety reports (ICSRs) after the market introduction of coronavirus disease 2019 (COVID-19) vaccines. Evidence from scientific literature suggests that when there is an extreme increase in the number of ICSR recorded in spontaneous reporting databases (such as the VAERS), an accompanying increase in the number of disproportionality signals (sometimes referred to as ‘statistical alerts’) generated is expected.

Objectives The objective of this study was to develop a natural language processing (NLP)-based approach to optimize signal management by excluding disproportionality signals related to listed adverse events following immunization (AEFIs). COVID-19 vaccines were used as a proof-of-concept.

Methods The VAERS was used as a data source, and the Finding Associated Concepts with Text Analysis (FACTA+) was used to extract signs and symptoms of listed AEFIs from MEDLINE for COVID-19 vaccines. Disproportionality analyses were conducted according to guidelines and recommendations provided by the US Centers for Disease Control and Prevention. By using signs and symptoms of listed AEFIs, we computed the proportion of disproportionality signals dismissed for COVID-19 vaccines using this approach. Nine NLP techniques, including Generative Pre-Trained Transformer 3.5 (GPT-3.5), were used to automatically retrieve Medical Dictionary for Regulatory Activities Preferred Terms (MedDRA PTs) from signs and symptoms extracted from FACTA+.

Results Overall, 17% of disproportionality signals for COVID-19 vaccines were dismissed as they reported signs and symptoms of listed AEFIs. Eight of nine NLP techniques used to automatically retrieve MedDRA PTs from signs and symptoms extracted from FACTA+ showed suboptimal performance. GPT-3.5 achieved an accuracy of 78% in correctly assigning MedDRA PTs.

Conclusion Our approach reduced the need for manual exclusion of disproportionality signals related to listed AEFIs and may lead to better optimization of time and resources in signal management.

1 Introduction

The Vaccine Adverse Event Reporting System (VAERS) has served as the cornerstone in the surveillance of adverse events following immunization (AEFIs) during the pandemic, playing an important role in detecting the initial signals of thrombosis with thrombocytopenia syndrome and myocarditis [1]. Recently, the US Centers for Disease Control and Prevention (CDC) reported that additional primary doses and booster coronavirus disease 2019 (COVID-19)

vaccine doses offer higher protection against severe disease with the Omicron variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). If expanded coverage of additional primary and booster doses is recommended, detecting AEFIs is crucial [2], as indeed, in general, for ongoing surveillance of vaccines effects in real-world ongoing immunization programs.

However, VAERS has already been challenged by an extreme increase in the number of individual case safety reports (ICSRs) after the market introduction of COVID-19 vaccines, which is also related to an extensive COVID-19 vaccination campaign in the United States (US) [3] and reporting bias, stemming from widespread media coverage

Extended author information available on the last page of the article

Key Points

This study aimed to optimize signal management in Vaccine Adverse Event Reporting Systems (VAERS) by developing a natural language processing (NLP) approach. By using NLP techniques, the study sought to automatically exclude disproportionality signals related to known adverse events following coronavirus disease 2019 (COVID-19) vaccination, streamlining the identification and removal of these signals.

The results showed that the NLP-based approach successfully dismissed 17% of disproportionality signals for COVID-19 vaccines, as they aligned with listed adverse events. Among the NLP techniques evaluated, Generative Pre-Trained Transformer 3.5 (GPT-3.5) demonstrated an accuracy of 78% in assigning Medical Dictionary for Regulatory Activities preferred terms. This approach reduces the need for manual exclusion of signals, leading to better time and resource management in signal analysis.

of vaccines during the pandemic [4] [Online Resource 1]. Evidence from scientific literature suggests that when there is an extreme increase in the number of ICSRs recorded in spontaneous reporting databases (such as the VAERS), an accompanying increase in the number of disproportionality signals (sometimes referred to as ‘statistical alerts’) generated is expected [5]. The first part of processing disproportionality signals involves manual exclusion of already listed AEFIs. This is a time- and resource-intensive task and tools that can help automate this process are urgently needed [6]. In this study, we therefore aimed to automate this exclusion of already known AEFIs using the Finding Associated Concepts with Text Analysis (FACTA+), after first reviewing the literature to ensure the novelty of our application in the context of signal management (for more detail, see Online Resource 2). We used VAERS data on COVID-19 vaccines as a proof-of-concept to test the performance of our approach.

2 Methods

2.1 Data Source

The VAERS is a spontaneous reporting database created by the US Food and Drug Administration (FDA) and CDC to monitor the safety of vaccines after immunization. The VAERS is a part of passive surveillance and collects reports of suspected AEFIs from different sources, which include healthcare providers and vaccine manufacturers. Patients, parents, and caregivers are also encouraged to report AEFIs

to the VAERS. AEFIs recorded in the VAERS are coded using the Medical Dictionary for Regulatory Activities (MedDRA). VAERS data are publicly available. In this study, we collected data from the VAERS during the period 2020–2022 [7].

2.2 Disproportionality Analysis

Disproportionality analysis was conducted according to the guidelines from the CDC [5, 8]. A contingency table was used to calculate the proportional reporting ratio (PRR): $a/(a+b)/c/(c+d)$ [Table 1]. If the lower boundary of the 95% confidence interval (CI) was > 1 , the threshold for considering the vaccine-AEFI pair as a disproportionality signal was reached. The 95% CI of the PRR is calculated as follows:

$$PRR \pm e^{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

We have chosen to conduct the analysis for all COVID-19 vaccines simultaneously, rather than stratifying by individual vaccines. This decision stems from the understanding that stratification by vaccine type would significantly reduce the overall number of disproportionality signals generated. Such stratification would limit the scope of our analysis and, in turn, curtail our ability to thoroughly assess the effectiveness of our method in identifying and screening out signals related to listed AEFIs when a lot of disproportionality signals are generated. Our primary objective was to subject our method to the most challenging scenario, involving an extreme volume of disproportionality signals, to rigorously test its capacity to distinguish and filter out signals associated with listed AEFIs. By conducting a unified analysis across all COVID-19 vaccines, we believe we have created the optimal conditions to rigorously evaluate the performance of our method under these extreme conditions.

2.3 Dismissal of Disproportionality Signals Related to Listed Adverse Events Following Immunization

FACTA+ was used to retrieve related signs and symptoms of listed AEFIs from MEDLINE. Listed AEFIs were retrieved from the Summary of the Product Characteristics of COVID-19 vaccines (Online Resource 3). FACTA+ is an advanced text-mining tool that helps discover associations

Table 1 Contingency table

	AEFIs of interest	Other AEFIs
COVID-19 vaccines	a	b
Other vaccines	c	d

AEFIs adverse events following immunization, *COVID-19* coronavirus disease 2019

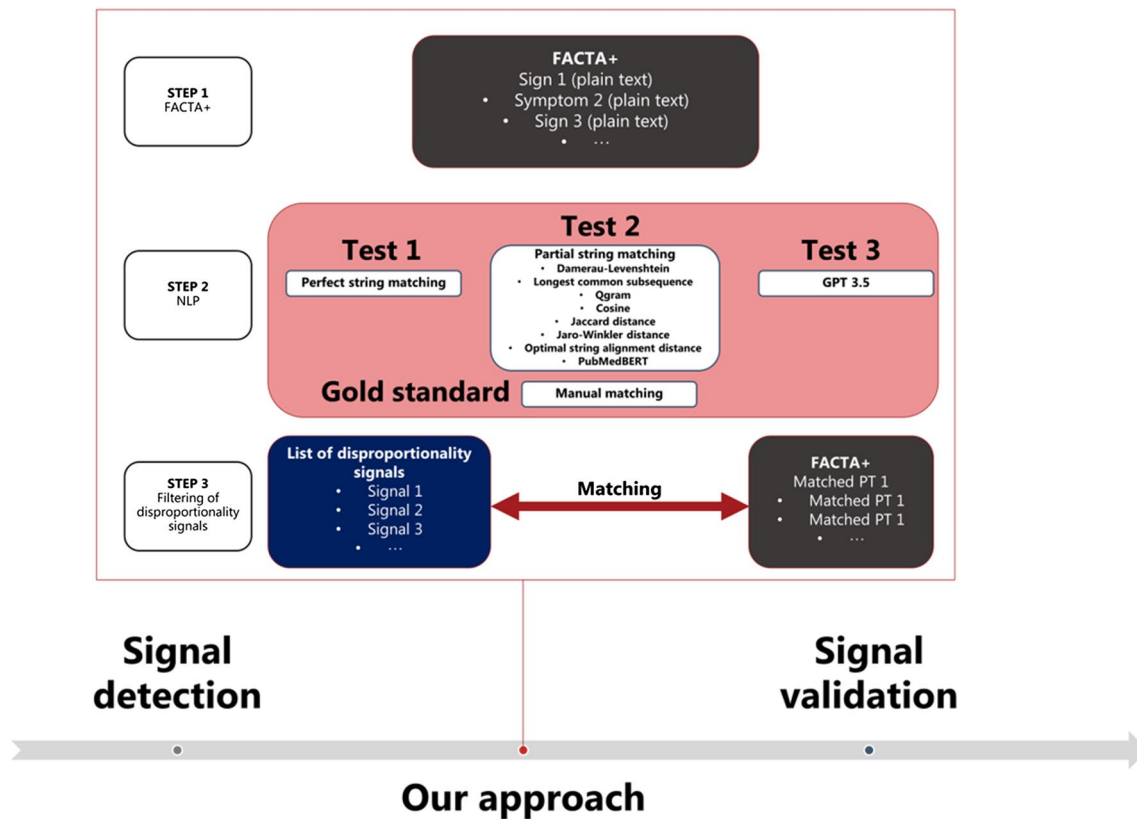


Fig. 1 Our proposed approach in the first steps of signal management. *Step 1:* Using FACTA+ to retrieve biomedical concepts ('signs and symptoms') associated with the list of AEFIs from VAERS for the COVID-19 vaccines. *Step 2:* Converting the signs and symptoms obtained from FACTA+ into MedDRA PTs using either perfect string matching, partial string matching, or large-language model embedding similarity. *Step 3:* Excluding disproportionality signals associ-

ated with signs and symptoms of labeled AEFIs. *FACTA+* Finding Associated Concepts with Text Analysis, *AEFIs* adverse events following immunization. *VAERS* Vaccine Adverse Event Reporting System, *COVID-19* coronavirus disease 2019, *MedDRA* Medical Dictionary for Regulatory Activities, *PT* preferred terms, *NLP* natural language processing, *GPT 3.5* Generative Pre-Trained Transformer 3.5

between biomedical concepts from MEDLINE articles. The whole MEDLINE corpus containing more than 20 million articles is indexed with an efficient text search engine. In FACTA+, a broad range of important biomedical concepts is covered by the combination of a machine learning-based term recognizer and large-scale dictionaries for genes, proteins, diseases, and chemical compounds [9].

From FACTA+, signs and symptoms of listed AEFIs were extracted as plain texts and were not immediately usable for identifying disproportionality signals to be dismissed (Online Resource 4). Consequently, we had to convert plain texts from FACTA+ to MedDRA preferred terms (PTs). To identify MedDRA PTs of signs and symptoms of listed AEFIs, we screened MedDRA PTs manually and using natural language processing (NLP) techniques. We explored NLP techniques to investigate whether it was feasible to automate the identification of MedDRA PTs from plain text obtained from FACTA+ (Fig. 1).

Nine traditional NLP techniques were examined. Specifically, perfect string matching and seven string distance metrics—Damerau–Levenshtein, longest common subsequence, qgram, cosine, Jaccard distance, Jaro–Winkler distance, and optimal string alignment distance [10], as well as semantic representations generated by pretrained language models. For this latter approach, we used PubMedBERT [11] with the SentenceTransformer library to create sentence embeddings of the extracted signs and symptoms as well as MedDRA ontology [12], where each sentence embedding is a mean embedding of the encoder output. The most similar term to a sign/symptom is then identified using cosine distance.

We then employed the aforementioned distance metrics in our NLP techniques as follows:

- (1) We utilized the complete MedDRA version 23 to incorporate a proxy for accurately matched PTs.

- (2) For each NLP technique, we computed density functions of the metrics separately for good and bad matches. In this context, good matches were determined through manual assessment by the lead author. Specifically, we considered a match to be good when the matched term was identical to the assessor's preferred term in MedDRA.
- (3) By calculating the overlapping area of the density functions, we chose a cut-off of 30% of the overlapping area to discern whether a clear separation existed between the density functions of good and bad matches. This decision was made because if the overlapping area of good and bad matches is too extensive, it means that we cannot use the aforementioned partial string match metrics to effectively differentiate between good and bad matches, as their values are very similar in both cases. Opting for a higher threshold, nearing 50%, would have made it nearly indistinguishable from random guessing to determine whether a match was performed correctly or poorly. The choice of a 30% threshold is non-conservative, as it allows for the possibility of other partial string match measures being employed. In fact, opting for a highly conservative cut-off, such as 5%, where only 5% of the values between good and bad matches overlap, would have unfairly penalized other partial string matching methods compared with Generative Pre-Trained Transformer (GPT-3.5), as we did not anticipate any of them performing at such high levels. Instead, we allowed for some margin of error to evaluate their feasibility, but unfortunately they did not demonstrate satisfactory performance.
- (4) If the overlapping area was below 30%, we proceeded to compute the accuracy of matching using manual matches as the gold standard. However, if none of the techniques exhibited an overlapping area below 30%, we disregarded the use of NLP techniques.

Additionally, we used GPT-3.5. We used this state-of-the-art large language model to facilitate the retrieval of MedDRA terms from unstructured text extracted from FACTA+ without relying on distance metrics. We performed this specific task using the OpenAI API. We iterated over each sign and symptom retrieved from FACTA+ a POST request to the OpenAI API endpoint (<https://api.openai.com/v1/chat/completions>). The request included certain headers, such as the API key for authorization and the content type as JSON. The body of the request contains information about the model to be used (in this case, 'gpt-3.5-turbo-16k') and a message from the user. The message was generated dynamically, based on the signs and symptoms extracted from FACTA+.

It asked for the PT in the MedDRA that was the most similar in meaning, and the string format to a specific sign and symptom from FACTA+. We repeated this process for each sign and symptom retrieved from FACTA+. Specifically, we used the following prompt: *"Output only the preferred term in MedDRA that has the most semantic and string similarity to (Sign or Symptom from FACTA+)"* (Accessed March 2023). To assess the consistency of the answers generated by GPT-3.5, we conducted the same analysis 10 times. Each analysis involved inputting the same data and posing the same question to the model. By performing multiple iterations of the analysis, we aimed to evaluate the stability and reliability of the responses generated by GPT-3.5. Through this process, we assessed whether the model consistently provided similar or divergent answers across the different runs. Cohen's kappa (κ) was used to assess the stability and reliability of the answers. To further evaluate the differences in the answers generated by GPT-3.5 during the repeated analysis, we employed Cohen's kappa (κ) statistic. Cohen's kappa (κ) is a measure of inter-rater agreement that assesses the level of agreement beyond what would be expected by chance alone. By calculating Cohen's kappa (κ) for the answers provided in each analysis iteration, we aimed to quantify the degree of consistency in GPT-3.5's responses. This statistical measure helped us determine whether the variations observed across the multiple runs were within an acceptable range of agreement or if there were substantial differences among the generated answers. The use of Cohen's kappa (κ) provided a standardized metric to assess the consistency of GPT-3.5's outputs, and aided in quantifying the level of concordance among the repeated analyses.

For those NLP techniques with an overlapping area < 30%, and for GPT-3.5, we computed the accuracy or rather the proportion of good match from the plain text from FACTA+. We determined the most effective approach for matching signs and symptoms retrieved from FACTA+ with MedDRA PTs by prioritizing accuracy. The following steps were taken:

- (1) We first evaluated the accuracy of the perfect string match algorithm.
- (2) Next, we assessed the accuracy of NLP techniques using partial string matching and with an overlapping area that was < 30.
- (3) Additionally, we evaluated the accuracy of GPT-3.5.

If any of the NLP techniques achieved 100% accuracy, we utilized it. Otherwise, we manually matched the MedDRA PTs for signs and symptoms of listed AEFIs as they were considered to have 100% accuracy. We then checked

if the selected method matched any of the PTs of disproportionality signals for COVID-19 vaccines. We computed the frequency and proportion of successful matches and identified the disproportionality signals that were dismissed using this approach.

2.4 Dismissed Signals in the Important Medical Event (IME) List

Challenges with the use of hierarchical terminologies is one of the great rate limiters of signal detection in spontaneous reports [13]. While an AEFI may have already been listed, if a listed AEFI exhibits symptoms indicating a more severe outcome, it constitutes a new signal. This scenario can arise despite the prior labeled status. In the context of this study, the latter scenario poses a potential problem. Therefore, we performed an additional analysis to investigate whether our approach filters out symptoms or sentiments indicative of worsening outcomes.

Therefore, we assessed how the signs and symptoms reported in the ICSR and within the IME list [14] that were identified as disproportionality signals changed over the years. We investigated and plotted the disproportionality signals within the IME list that were not common in 2020, 2021, and 2022. We then assessed whether these signals suggested a worsening outcome of a labeled AEFI by plenary discussion among two authors (MS and GD).

2.5 Finding Associated Concepts with Text Analysis (FACTA+)

FACTA+ is a text-mining tool for exploring associations between biomedical concepts trained on the extensive MEDLINE article database, comprising over 20 million articles [9, 15]. This innovative system employs a robust text search engine for efficient indexing of the entire MEDLINE corpus, enabling interactive exploration of associations between concepts and their supporting textual evidence. Users can input arbitrary query terms, and FACTA+ returns relevant concepts. It encompasses extensive dictionaries for 80,260 genes, proteins, diseases, and chemical compounds. FACTA+ generates a ranked list of target concepts based on their ‘expected information’, which is a measure of how often the query term and a target concept occurs compared with what would be expected if they were unrelated.

3 Results

During the study period, 757 disproportionality signals were generated using quantitative methods as used routinely in the field and as listed in CDC guidance.

3.1 Natural Language Processing to Automate Translation from Plain Text to Medical Dictionary for Regulatory Activities Preferred Terms

On the set of 906 signs and symptoms that could manually be mapped to MedDRA, only 56% of signs and symptoms extracted from FACTA+ as plain text was correctly converted into MedDRA PTs by perfect string matching (accuracy: 56%).

None of the NLP techniques using partial string matching delivered a clear separation boundary that could help to separate good from bad matches, as we observed a minimum overlap of density functions of 65% (Fig. 2).

The accuracy of GPT-3.5 was 78%, which means that 8/10 MedDRA PTs were correctly assigned. In the results of our analysis using GPT-3.5, we observed perfect agreement among all 10 attempts, as indicated by Cohen's kappa (κ) statistic. The calculated kappa value of 1.0 suggested complete concordance in the answers generated by the model across the repeated analyses.

We therefore had to rely on the manual matching of signs and symptoms retrieved from FACTA+ with PTs from MedDRA.

3.2 Dismissal of Signals

By using manually matched PTs, our approach dismissed 130/757 (17%) disproportionality signals (Online Resource 5).

3.3 Dismissed Signals within IMEs

In total, 399 disproportionality signals for IMEs were not common in 2020, 2021, and 2022. Twenty-one of the 399 disproportionality signals were dismissed by our approach, and none suggested an increase in severity for listed AEFIs (Fig. 3; Table 2).

4 Discussion

Signal management is critical to vaccine vigilance, ensuring that AEFIs are appropriately identified and evaluated [16]. The detection and analysis of disproportionality signals in adverse event reporting systems can be a challenging task, particularly when dealing with a large volume of data. Signals of disproportionality must undergo clinical review before they are considered signals of suspected causality. This is a time-consuming task, and any advances to better focus this clinical review can provide potentially enormous benefits [17–22]. In this study, we employed a combination of manual and automated techniques to dismiss signals

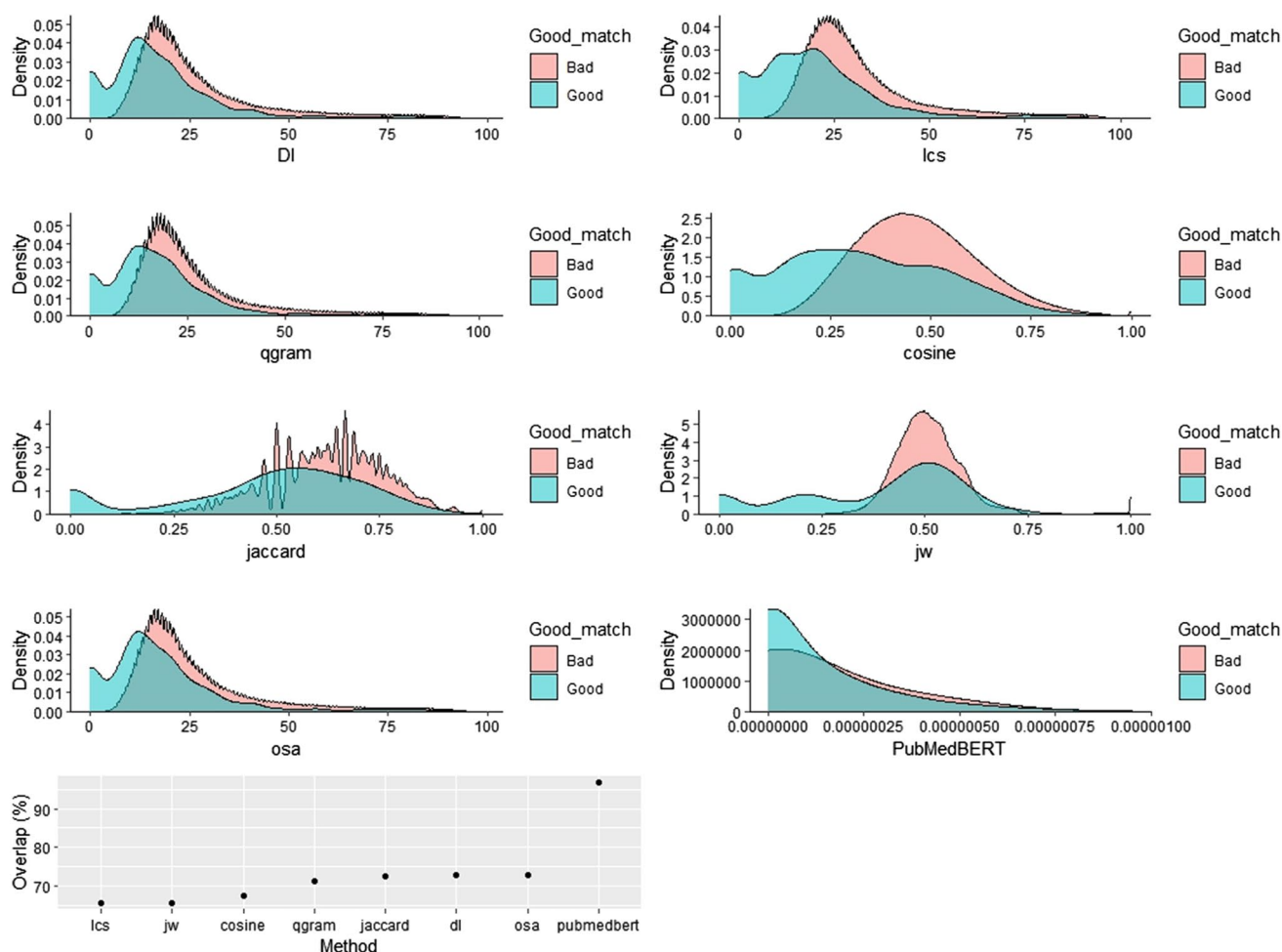


Fig. 2 Partial string matching of signs and symptoms from FACTA+ with Medical Dictionary for Regulatory Activities preferred terms using eight NLP techniques. *DL* Damerau–Levenshtein, *LCS* longest

common subsequence, *JW* Jaro–Winkler distance, *OSA* optimal string alignment distance, *FACTA+* Finding Associated Concepts with Text Analysis

Fig. 3 Number of disproportionality signals that were not in common in 2020, 2021, and 2022. *IME* important medical event

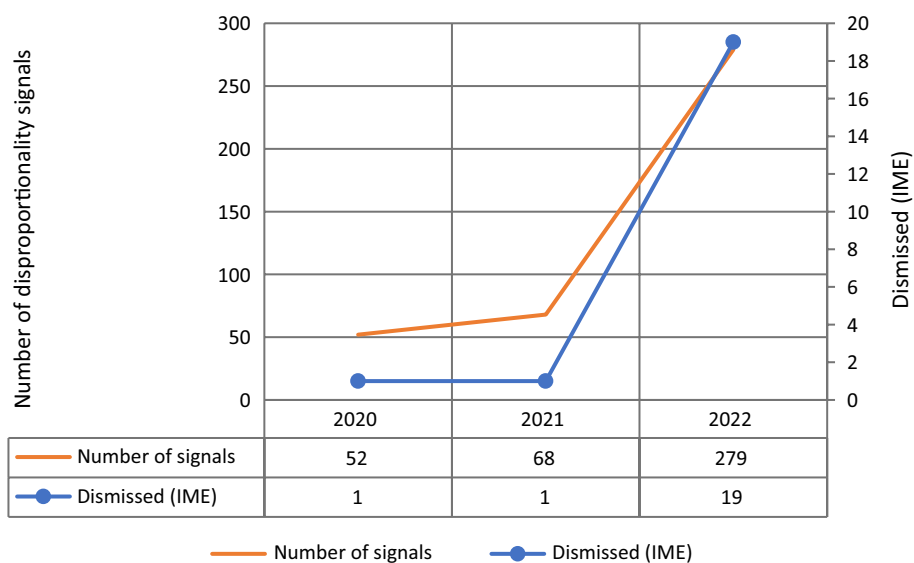


Table 2 Signals dismissed in the IME list

Signals	Year
Arthritis	2022
Arthropathy	2022
White blood cell count increased	2022
Monocyte count increased	2022
Chronic kidney disease	2022
Deafness unilateral	2022
Dementia	2022
Gastrointestinal hemorrhage	2022
Hematemesis	2022
Hematuria	2022
Hemiparesis	2022
Hyperkalemia	2022
Bradycardia	2022
Postmenopausal hemorrhage	2022
Respiratory distress	2022
Bronchitis	2022
Shock	2022
Neutrophil count increased	2022
Anaphylactic reaction	2022
Hallucination	2021
Angioedema	2020

IME important medical event

of disproportionality, and utilized NLP for the translation of plain-text signs and symptoms of listed AEFIs retrieved from the scientific literature into MedDRA PTs. Of note, to reduce this burden, researchers have explored various approaches to automate the signal management process. Data-mining techniques, such as data clustering, association rule mining, and machine learning algorithms, have been utilized to identify patterns and relationships within the adverse event data, enabling more efficient and targeted signal detection [23, 24].

The results of our statistical alert dismissal process were promising, as our approach successfully dismissed 17% of disproportionality signals. By accurately eliminating irrelevant signals, resources can be allocated more efficiently towards investigating potential adverse events that warrant further analysis.

Regarding our attempt to automate the translation of plain-text symptoms to MedDRA PTs using NLP techniques, this presented several challenges. When we employed basic NLP techniques [10], we found that only 56% of signs and symptoms extracted from FACTA+ perfectly matched with MedDRA PTs. None of the NLP methods using partial string matching demonstrated a clear separation boundary that could effectively distinguish good matches from poor

matches. This was evident from the density functions, which exhibited a minimum overlap of 65%.

The obtained result of 78% accuracy for GPT-3.5 is noteworthy. The high accuracy suggests that GPT-3.5 possesses a significant level of apparent understanding and proficiency in identifying and categorizing MedDRA PTs. These findings are promising, as accurate PT assignment is crucial for efficient and effective pharmacovigilance, adverse event reporting, and medical documentation. However, it is important to acknowledge the remaining 22% of PTs that were not correctly assigned, as there is still room for improvement. Further research and enhancements to GPT-3.5's training and fine-tuning processes could potentially increase its accuracy and make it an even more valuable tool in the field of signal management. The high level of agreement across 10 attempts of PT assignments suggests strong consistency in GPT-3.5's responses, and indicates a robust and reliable performance of the model for the specific task at hand. Despite the exciting results from GPT-3.5, manual matching of signs and symptoms from FACTA+ with lower-level terms (LLTs) from MedDRA was deemed necessary to ensure accurate translation of plain-text signs and symptoms into the appropriate MedDRA PTs. Manual matching of signs and symptoms retrieved from FACTA+ (or any other system) to PTs can be a labor-intensive and time-consuming task. It involves the manual alignment of concepts and requires human experts to review and match the output of FACTA+ to the appropriate PTs. This process is prone to human error and subjectivity, which may introduce inconsistencies and affect the accuracy of the results.

Therefore, at this stage, despite the progress in NLP and other techniques, this study emphasizes the continued importance of human expertise in pharmacovigilance. However, in our study, the combination of automated methods with manual verification allowed a more reliable and accurate signal validation.

It is interesting to observe that the results obtained when using different NLP techniques were different. This result highlights the impact that algorithmic differences, sensitivity to textual differences, string length and complexity, and language and domain specificity have on performing this task [25].

Regarding severity of signs and symptoms detected as disproportionality signals, we investigated whether the dismissed disproportionality signals were present in the IME list over the years to investigate whether we were dismissing disproportionality signals suggestive of an increased severity of labeled AEFIs [14]. Among those dismissed in the IME list, none suggested an increase in the severity of listed AEFIs.

Lastly, while we have not investigated the use of resources such as MedNorm [26] or PubMedBERT for classification

tasks, one such avenue could involve incorporating resources such as MedNorm [27] to explicitly train PubMedBERT for multiclass classification of MedDRA terms. Approaches such as XTARS [28] methodology could potentially enhance the accuracy and robustness of classification tasks. However, it is important to note that these approaches may also have their own limitations, such as the availability and quality of training data (29, 30).

4.1 Limitations

While our study has provided valuable insights into the use of FACTA+, it is important to acknowledge certain limitations that should be taken into consideration. First, our NLP approach focused on an unsupervised approach, which may have its own limitations in terms of precision and recall. Supervised approaches, on the other hand, could potentially provide more accurate results by leveraging labeled data for training NLP models. Second, our study focused on the application of string similarity metrics, which may not capture the semantic meaning of medical terms comprehensively. This could result in potential misclassifications or ambiguities in the classification process. Additionally, the use of FACTA+ itself may have inherent limitations, such as its reliance on specific resources or datasets, which could introduce biases or restrict its generalizability to other domains or languages. Overall, these limitations highlight the need for further research and exploration to address these concerns and improve the performance and applicability of FACTA+ in medical classification tasks.

Another potential limitation of this research study is the lack of specific information regarding the version of MedDRA used in the analysis. The NLP language model employed in the study, based on the GPT-3.5 architecture, was trained on a diverse range of data up until September 2021. Consequently, it does not possess explicit knowledge of the MedDRA version released after that time. As the MedDRA terminology is regularly updated to incorporate new medical knowledge and evolving terminology, it is essential to consider whether the version used in this research aligns with the most current standards. Future studies could benefit from incorporating the exact version of MedDRA utilized, thereby ensuring the accuracy and relevance of the findings within the context of the specific MedDRA version.

Regarding our analysis of the severity of signs and symptoms detected as disproportionality signals, it is important to mention that we only used 3 years of data and signals not common among those years, which certainly imposes limitations. One potential issue is the lack of granularity in the analysis, as a smaller sample size may not fully capture the range and variation of the data. Additionally, if the signals being analyzed are not consistent across the 3 years,

the analysis may not provide a comprehensive understanding of the underlying patterns and relationships in the data. Therefore, it is important to carefully consider the limitations when interpreting the results of our analysis.

4.2 Implications

The results of this study have significant implications for the field of pharmacovigilance and vaccine vigilance. First, the successful combination of the manual and automated signal dismissal approach provides a valuable way to improve the efficiency and accuracy of the signal validation process.

The challenges observed in converting plain-text symptoms to MedDRA PTs using NLP algorithms, on the other hand, demonstrate the limitations of the current automated methods. While NLP techniques show promise, there is a need for further research and development to enhance their accuracy and effectiveness in this context. Improving the performance of NLP algorithms to accurately convert plain-text symptoms into MedDRA PTs would considerably streamline the signal validation process and lessen reliance on manual matching. Further research should focus on exploring innovative approaches and technologies, such as deep learning models, which may provide further insights into improving automated signal dismissal and NLP translation.

Moreover, it is essential to continue evaluating the dismissed signals within the IME list and their implications for the severity of listed AEFIs. Monitoring and analyzing the trends and patterns of dismissed signals can help identify emerging risks or potential safety concerns that may have been overlooked. Long-term studies assessing the temporal patterns and severity changes in listed AEFIs should be conducted to ensure the continued safety and effectiveness of immunization programs.

Finally, we firmly believe that the methodology we have developed possesses a high degree of transferability to other vaccines/drugs and spontaneous reporting databases. This confidence stems from the foundational principles that underpin these databases and the striking similarities observed in their data structures. The core principles governing the collection and organization of adverse event reports remain consistent across various databases within the same domain. Consequently, we are convinced that our approach, which has demonstrated its efficacy in our current dataset, can be extended to analyze and extract valuable insights from other spontaneous reporting databases. This belief underscores our commitment to providing a robust and broadly applicable solution for the broader scientific community and stakeholders who rely on these databases for pharmacovigilance and healthcare decision making.

5 Conclusion

We tested a new NLP-based approach to automatically exclude disproportionality signals related to listed AEFIs by using signs and symptoms of listed AEFIs extracted from the scientific literature. This strategy may lead to better optimization of time and resources in signal validation.

Additionally, this study emphasizes the potential of combining manual and automated techniques in pharmacovigilance signal validation while highlighting the need for further research to improve NLP techniques and enhance automation. Continued efforts to refine automated methods, alongside human expertise, will contribute to more accurate and efficient pharmacovigilance practices, ultimately enhancing patient safety.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40264-023-01381-6>.

Funding Open access funding provided by Royal Library, Copenhagen University Library.

Declarations

Funding This study was not supported by any funding.

Conflict of interest Andrew Bate and François Haguinet are employees of GSK and hold stock and stock options. Guojun Dong, Gabriel Westman, Luise Dürlich, Anders Hviid, and Maurizio Sessa have no potential conflicts of interest to declare.

Ethics approval Not applicable.

Consent to participate Not applicable

Consent for publication Not applicable.

Availability of data and materials We used publicly available data from the VAERS that can be accessed at <https://vaers.hhs.gov/data/datasets.html>.

Code availability Can be requested upon reasonable request to the corresponding author.

Author contributions MS had the idea for the project. All authors drafted and commented on the manuscript. GJ, MS, and LD performed data management and analysis. All authors read and approved the final version of this manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory

regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Arepally GM, Ortel TL. Vaccine-induced immune thrombotic thrombocytopenia: what we know and do not know. *Blood*. 2021;138(4):293–8.
2. Prasad N, Derado G, Nanduri SA, Reses HE, Dubendris H, Wong E, et al. Effectiveness of a COVID-19 additional primary or booster vaccine dose in preventing SARS-CoV-2 infection among nursing home residents during widespread circulation of the omicron variant—United States, February 14–March 27, 2022. *MMWR Morb Mortal Wkly Rep*. 2022;71(18):633–7.
3. Ceacareanu A, Wintrob ZP. Summary of COVID-19 vaccine-related reports in the vaccine adverse event reporting system. *J Res Pharm Pract*. 2021;10(3):107.
4. Spiteri J. Media bias exposure and the incidence of COVID-19 in the USA. *BMJ Glob Health*. 2021;6(9): e006798.
5. Vaccine Adverse Event Reporting System (VAERS) standard operating procedures for COVID-19 Centers for Disease Control and Prevention. 2022 [cited 10 Oct 2022]. Available at: <https://www.cdc.gov/vaccinesafety/pdf/VAERS-COVID19-SOP-02-02-2022-508.pdf>.
6. Grundmark B, Holmberg L, Garmo H, Zethelius B. Reducing the noise in signal detection of adverse drug reactions by standardizing the background: a pilot study on analyses of proportional reporting ratios-by-therapeutic area. *Eur J Clin Pharmacol*. 2014;70(5):627–35.
7. Chen RT, Rastogi SC, Mullen JR, Hayes SW, Cochi SL, Donlon JA, et al. The Vaccine Adverse Event Reporting System (VAERS). *Vaccine*. 1994;12(6):542–50.
8. Faillie JL. Case-non-case studies: principle, methods, bias and interpretation. *Therapie*. 2019;74(2):225–32.
9. Tsuruoka Y, Miwa M, Hamamoto K, Tsujii J, Ananiadou S. Discovering and visualizing indirect associations between biomedical concepts. *Bioinformatics*. 2011;27(13):i111–9.
10. Lu J, Lin C, Wang W, Li C, Wang H (eds). String similarity measures and joins with synonyms. In: *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. New York: ACM; 2013.
11. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. *arXivorg*. 2021.
12. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-Networks. *arXivorg*. 2019.
13. Bate A, Brown EG, Goldman SA, Hauben M. Terminological challenges in safety surveillance. *Drug Saf*. 2012;35(1):79–84.
14. Inclusion/exclusion criteria for the “Important Medical Events” list European Medicines Agency 2021 [cited 8 May 2023]. Available at: https://www.ema.europa.eu/en/documents/other/inclusion-exclusion-criteria-important-medical-events-list-meddra_en.pdf.
15. Tsuruoka Y, Tsujii J, Ananiadou S. FACTA: a text search engine for finding associated biomedical concepts. *Bioinformatics*. 2008;24(21):2559–60.
16. Definition and application of terms for vaccine pharmacovigilance Council for International Organizations of Medical Sciences (CIOMS) 2012 [cited 24 Apr 2023]. Available at: <https://cioms.org>.

[ch/wp-content/uploads/2017/01/report_working_group_on_vaccine_LR.pdf](https://www.eurimdr.org/wp-content/uploads/2017/01/report_working_group_on_vaccine_LR.pdf).

17. Sessa M, Mascolo A, Callreus T, Capuano A, Rossi F, Andersen M. Direct-acting oral anticoagulants (DOACs) in pregnancy: new insight from VigiBase®. *Sci Rep*. 2019;9(1):7236.
18. Sessa M, Rafaniello C, Sportiello L, Mascolo A, Scavone C, Mac-cariello A, et al. Campania Region (Italy) spontaneous reporting system and preventability assessment through a case-by-case approach: a pilot study on psychotropic drugs. *Expert Opin Drug Saf*. 2016;15(Suppl 2):9–15.
19. Sessa M, Sportiello L, Mascolo A, Scavone C, Gallipoli S, di Mauro G, et al. Campania Preventability Assessment Committee (Italy): a focus on the preventability of non-steroidal anti-inflammatory drugs' adverse drug reactions. *Front Pharmacol*. 2017;8:305.
20. Sessa M, Rossi C, Mascolo A, Grassi E, Fiorentino S, Scavone C, et al. Suspected adverse reactions to contrast media in Campania Region (Italy): results from 14 years of post-marketing surveillance. *Expert Opin Drug Saf*. 2015;14(9):1341–51.
21. Sessa M, Rossi C, Rafaniello C, Mascolo A, Cimmaruta D, Scavone C, et al. Campania preventability assessment committee: a focus on the preventability of the contrast media adverse drug reactions. *Expert Opin Drug Saf*. 2016;15(Suppl 2):51–9.
22. Sessa M, di Mauro G, Mascolo A, Rafaniello C, Sportiello L, Scavone C, et al. Pillars and pitfalls of the new pharmacovigilance legislation: consequences for the identification of adverse drug reactions deriving from abuse, misuse, overdose, occupational exposure, and medication errors. *Front Pharmacol*. 2018;9:611.
23. Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, et al. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. *J Am Med Inform Assoc*. 2013;20(1):184–92.
24. Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. *J Biomed Inform*. 2017;66:72–81.
25. Navarro G. A guided tour to approximate string matching. *ACM Comput Surv*. 2001;33(1):31–88.
26. Belousov M, Dixon WG, Nenadic G (eds). *MedNorm: a corpus and embeddings for cross-terminology medical concept normalisation*. In: *Proceedings of the Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task*. Stroudsburg: Association for Computational Linguistics; 2019.
27. Cho H, Choi W, Lee H. A method for named entity normalization in biomedical articles: application to diseases and plants. *BMC Bioinform*. 2017;18(1):451.
28. Wang P, Wang S, Huang R, Huang Z. Quantifying bounds of model gap for synchronous generators. *arXiv pre-print server*. 2021.
29. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf*. 2014;37(10):777–90.
30. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc*. 2014;21(2):221–30.

Authors and Affiliations

Guojun Dong¹ · Andrew Bate^{2,3} · François Haguinet² · Gabriel Westman⁴ · Luise Dürlich^{5,6} · Anders Hviid^{1,7} · Maurizio Sessa¹

✉ Maurizio Sessa
maurizio.sessa@sund.ku.dk

¹ Department of Drug Design and Pharmacology, University of Copenhagen, Jagtvej 160, 2100 Copenhagen, Denmark

² Global Safety, GSK, Brentford, UK

³ Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

⁴ Department of Medical Sciences, Uppsala University, Uppsala, Sweden

⁵ Department of Linguistics and Philology, Uppsala University, Uppsala, Sweden

⁶ Department of Computer Science, RISE Research Institutes of Sweden, Kista, Sweden

⁷ Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark