

COVID Vaccine and Cardiovascular Risks: A Natural Language Analysis of Vaccine Adverse Event Reports

Yiqing Zhao
Department of Preventive Medicine
Northwestern University
Chicago, USA
yiqing.zhao@northwestern.edu

Michael G. Ison
Department of Medicine (Infectious
Disease)
Northwestern University
Chicago, USA
mgison@northwestern.edu

Yuan Luo (Corresponding Author)
Department of Preventive Medicine
Northwestern University
Chicago, USA
yuan.luo@northwestern.edu

Abstract—Adverse events (AEs) following COVID vaccination have been intensely monitored. In our study, we developed a natural language processing system to analyze data from a spontaneous reporting system - Vaccine Adverse Event Reporting System and detect signals of AEs following administration of COVID vaccines. Our system included several components to magnify novel and rare AEs, including 1) excluding COVID positive patients, 2) excluding sentences discussing disease history or family history, 3) standardizing symptom concepts into 30 major AEs, 4) using influenza vaccine recipients as control group when calculating reporting odds ratio. We identified several cardiovascular and inflammatory-related AEs that demonstrated high odds ratio. We demonstrated our system can serve as a complementary system to identify and monitor AEs outside of pre-defined outcomes routinely monitored by existing databases or projects.

Keywords—COVID vaccine, natural language processing, concept standardization, rare adverse event

I. INTRODUCTION

Since the implementation of vaccination, adverse events (AEs) following vaccination have been intensely monitored [1]. The Centers for Disease Control and Prevention (CDC) has several well-established vaccine safety surveillance systems that monitor AEs for vaccines: 1) Vaccine Adverse Event Reporting System (VAERS) [2], a spontaneous (or passive) reporting system that curated reports submitted by the vaccine recipients or their caregivers regarding AEs after all recently introduced routinely administered vaccines. The strength of VAERS is that it curated data at the national level, it's updated frequently and allows rapid detects of safety signals, it enables detection of rare adverse events, and the data is available to the public. However, the limitations of VAERS include: reporting bias, inconsistent data quality and completeness of information, lack of unvaccinated comparison group, not being designed to assess causality [3]. 2) Vaccine Safety Datalink (VSD) project [4], conducts near-real-time active surveillance for vaccine safety and has performed weekly surveillance for pre-specified outcomes [5, 6]. VSD currently has 9 participating integrated healthcare organizations representing a diverse population across the US and has curated data on over 12 million people per year regarding pre-specified outcomes, covariates, unvaccinated concurrent comparators, vaccinated concurrent comparators, self-controlled risk interval, and historical comparators [7]. It was designed to detect statistically significant associations and statistical signals. (values above specified statistical threshold). When a statistically significant association or signal occurs, assessment involves a series of checks and evaluations. Chart

confirmation of diagnoses to confirm or exclude cases as true incident cases is a key part of the statistical signal assessment. 3) V-safe project [8], a new active surveillance system for collecting near-real-time data from COVID vaccine recipients in the US through voluntarily self-enrolled health surveys starting from the day of the first dose of COVID vaccine through 12 months after the final dose. Participants were asked questions about local and systemic reactions (e.g., injection site pain, fatigue, headache). An initial summary from V-safe monitoring of COVID vaccine for pre-specified outcomes was published in April 2021 [9].

In our study, we chose to use VAERS database because surveillance based on AERS databases has been a cornerstone for the early detection of drug/vaccine safety [10]. A recent study utilized VAERS to summarize incidences of anaphylaxis after administration of the Pfizer-BioNTech and Moderna vaccines [11]. We were aware that signals in VAERS were not evaluated for causal relationship and statistical significance as [12]. However, compared to other active surveillance systems [4, 8, 13] with pre-specified outcomes to be monitored, VAERS offers opportunities for early warnings of novel or rare AEs in a timely manner [12] and this is the focus of our study.

Several studies have explored use of natural language processing (NLP) for analyzing VAERS reports. Botsis et al. [14] developed a system - Vaccine Adverse Event Text Miner (VaeTM) that extracts “outcome of interest” (diagnosis, cause of death, second level diagnosis), “time to onset” and “alternative explanations” (drug, medical and family history) from VAERS reports. Baer et al. compared VaeTM extracted data the expert's abstraction of the three variables from VAERS reports. Proportion of matches between VaeTM extraction and expert's abstraction results were: 93% for “outcome of interest”, 78% for “alternative explanation” and 54% for “time to onset”. Du et al. [15] developed a deep learning system for extraction of Guillain-Barré syndrome (GBS) related to influenza vaccine from VAERS reports. They extracted major entities related to GBS, including, investigation, nervous_AE, other_AE, procedure, social_circumstance, and temporal_expression.

In our study, we developed an NLP system by integrating existing NLP APIs to identify AE entities (AE concept mentions and time-to-onset) following the administration of COVID vaccines. We then standardized different expressions of AE mentions that refer to the same concept. Finally, we calculated reporting odds ratio (ROR) to detect signals of standardized AE occurrences. We chose to use ROR instead

of proportional reporting ratio (PRR) because PRR was a biased estimator which could be improved by applying the principles of a case-control study and reformulating as an odds ratio, ROR [16]. The ROR is the odds of a certain AE occurring with the vaccine of interest (i.e., COVID vaccine), compared to the odds of the same event occurring with all other vaccines (i.e., other vaccines) in the database [16]. We only counted events in the influenza vaccine instead of all other vaccines when calculating ROR because 1) influenza and COVID virus are both enveloped, single-stranded RNA viruses [17]. 2) they both affect the respiratory system. 3) many symptoms of COVID patients resemble those of influenza patients [18]. Therefore, we assume that using the influenza vaccine as a control group when calculating ROR may balance out symptoms as a sign of the vaccine taking effect. The design could potentially magnify signals for novel and rare AEs.

II. MATERIALS AND METHODS

A. Data

We retrieved 41217 (31143 from Pfizer/Moderna, 9975 from Johnson & Johnson, 99 unknown) COVID vaccine-related reports (12/15/2020-4/15/2021) and 44429 influenza vaccine-related reports (1/1/2017-12/14/2020) from VAERS. Influenza vaccine reports were selected with no time overlap with the COVID vaccine to avoid potential interactions between the two diseases and two vaccines. Also, the total number of influenza vaccine reports was set to match the total number of COVID vaccine reports. Additionally, we excluded patients who were already tested positive for COVID because symptoms from being positive for COVID might be a confounder for the detected AEs.

B. NLP Pipeline

Our NLP system was implemented in the following steps (see Fig. 1):

1) We split each report into individual sentences and then utilized MetaMap API [19] to extract disease/symptom concepts automatically from each sentence.

2) We used MedTagger API [20] to detect the context of each concept so that we were able to identify negated concepts, historical event concepts, and time-to-onset of the AEs.

3) We standardized symptom concepts to a standardized AE expression using both the synonym dictionary from MetaMap API as well as expert confirmation. Code for standardization is available at: https://github.com/yzhao5183/COVID_VACCINE_AE.

4) We kept 30 major AEs with total occurrence > 10 and calculated ROR using COVID vs influenza vaccine reports. For counts equal to zero, Haldane correction was applied.

III. RESULTS

Demographic and manufacturer information regarding the cohort was listed in Table 1. Females reported more cases than males. The average age for COVID vaccine AE reporters is younger in the adult population and is younger in the female population.

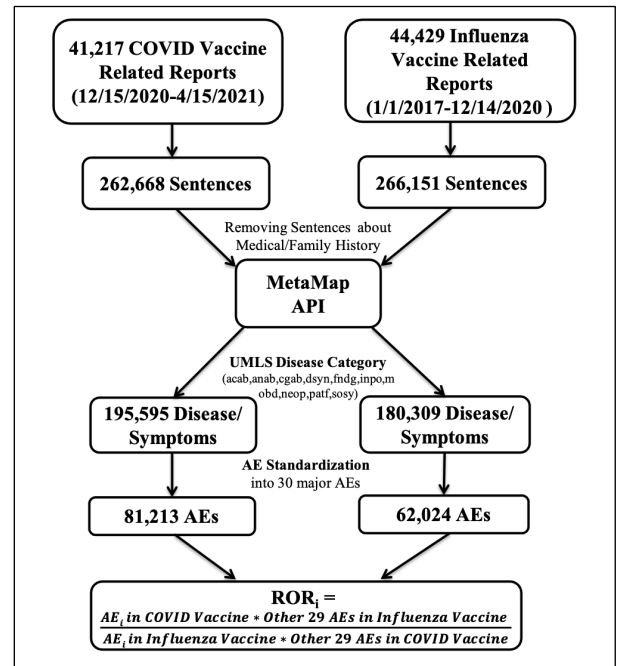


Fig. 1. NLP pipeline for AE signal detection.

TABLE I. DEMOGRAPHIC CHARACTERISTICS OF COHORTS FOR COVID AND INFLUENZA VACCINE AE REPORTS

	Manufacturer/ VaxType	Female	Male	Unknown	Grand Total
Count	PFIZER/MODERNA	21837	8477	829	31143
	JANSSEN	6601	3100	274	9975
	UNKNOWN COVID	55	38	6	99
	INFLUENZA	28243	13027	3159	44429
Avg Age ≥18	PFIZER/MODERNA	50.7	59.8	55.0	53.2
	JANSSEN	45.5	43.5	47.2	44.9
	UNKNOWN COVID	55.0	64.2	38.5	58.4
	INFLUENZA	55.9	58.1	57.9	56.5
Avg Age <18	PFIZER/MODERNA	14.9	15.7	14.1	15.0
	JANSSEN	16.3	16.4	16.6	16.3
	UNKNOWN COVID	17.0			17.0
	INFLUENZA	7.7	7.0	7.5	7.4

We identified 57764, 23265, and 62025 non-duplicate (removed duplicate occurrences identified in the same reports) AE occurrences from COVID (Pfizer/Moderna), COVID (J&J), and influenza vaccine reports respectively. Counts, time-to-onset (if available), and calculated RORs of each AE were listed in Fig 2, from which we found several AEs showing statistically significant higher RORs ($ROR > 3$, $p < 0.05$) for COVID vaccine recipients vs influenza vaccine recipients among people who have developed AEs and submitted AE reports: 1) Despite low incidence rate, acute appendicitis, cardiac attack/arrest, and hypoxia had significantly higher odds in COVID vaccine recipients compared to influenza vaccine recipients who developed AEs. 2) Notably, several acute cardiovascular events e.g., arrhythmia/palpitation, infarction/aneurysm/hemorrhage, tachycardia, and thrombus/embolism had higher odds in COVID vaccine recipients who developed AEs, which might be indicative of cardiovascular risks.

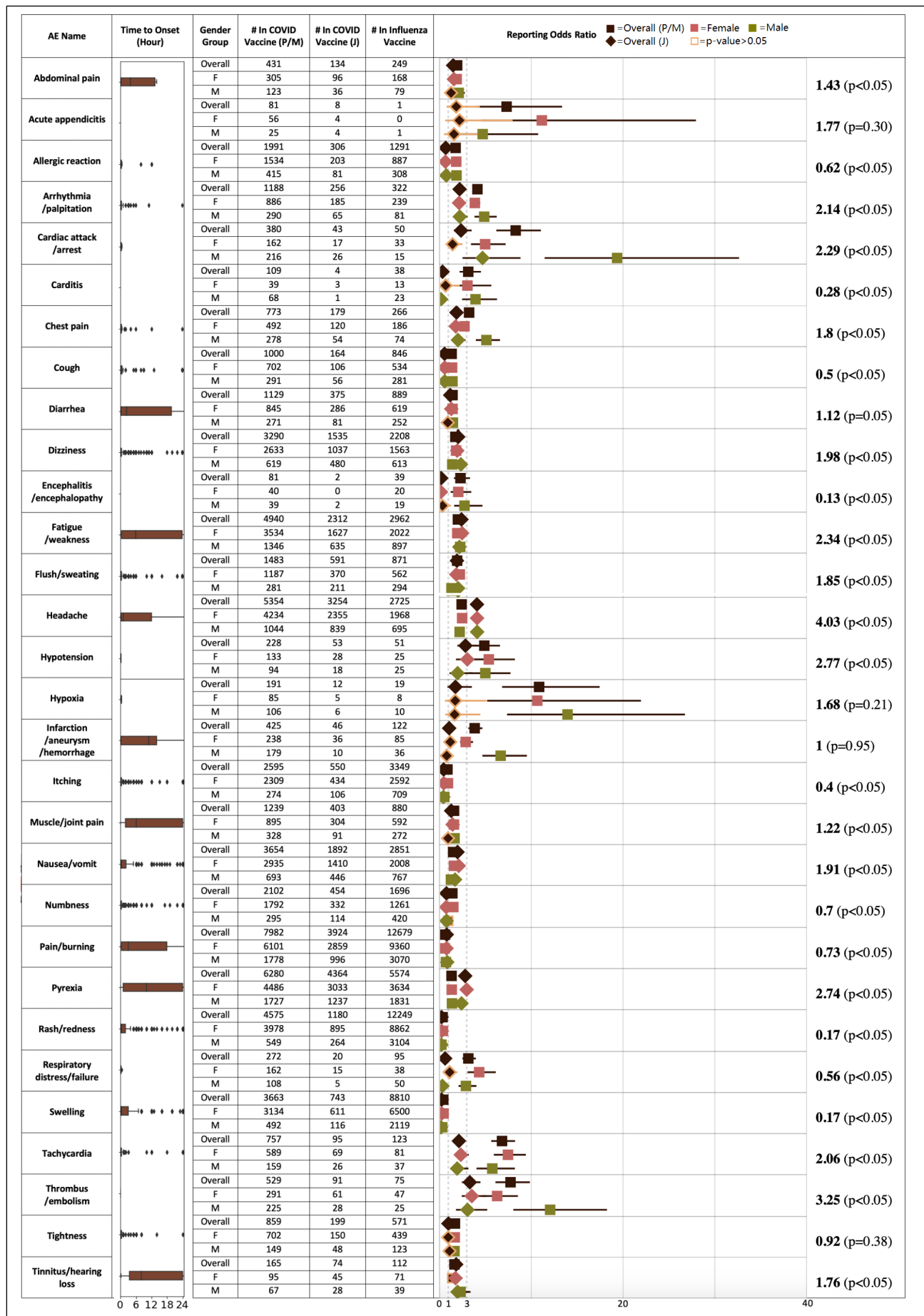


Fig. 2. Counts, Time-to-onset (in COVID vaccine) and RORs for COVID vs Influenza Vaccine-related AEs (Dashed Line is ROR=1 and 3; Listed ROR Value and P-value is Calculated using Overall Report Group).

While females contributed more AE reports, gender patterns in high ROR AEs were not conclusive. J&J's Janssen COVID-19 vaccine did not appear to have higher odds in most AEs. However, considering the number of people vaccinated by J&J (7.7 million doses as of 4/15/2021) versus Pfizer/Moderna (190.4 million doses as of 4/15/2021), incidence rates of AEs from the J&J vaccine were much higher. Onsets were retrieved only from COVID reports and we only plotted onset distribution within 24 hours. 13% of AEs were associated with an onset time within 24 hours. We could see that most severe AEs (e.g., arrhythmia, cardiac attach/arrest) were associated with a short onset, which might be indicative of an association with vaccination.

IV. DISCUSSION

We found COVID vaccine recipients had higher odds of developing cardiovascular and inflammatory (e.g., appendicitis, pancreatitis, carditis) AEs as compared to influenza vaccine recipients. The detected cardiovascular signals for carditis coincided with several recent reports of acute myocarditis following the COVID vaccine [21-28]. Additionally, the detected signals for hemorrhage and thrombus/embolism coincided with several recent reports of hemorrhage, blood clots, and thrombocytopenia following the administration of COVID vaccines in the UK [29]. The timely AE case reports and signal detection has already encouraged further investigation into the mechanisms of the novel and rare AE: several researchers hypothesized that antibodies stimulated by the COVID vaccine might trigger an autoimmune response to surface antigens on platelets or megakaryocytes, which could lead to the formation of blood clots and thrombocytopenia [29]. In the case for carditis signals, the Advisory Committee on Immunization Practices (ACIP) acted promptly after several case reports of myocarditis and pericarditis in mRNA COVID vaccine recipients, which predominantly occurred in young males after the second dose, and issued interim recommendations in June 2021. They determined the benefits of using mRNA COVID vaccines outweigh the after reviewing available evidence including that for risks of myocarditis. However, information on myocarditis after receipt of mRNA COVID vaccines was added to vaccine label to inform the public of the potential AE [30].

Despite higher RORs, the incidence rates of cardiovascular AEs were low considering the number of people vaccinated (125.8 million on 04/15/2021, at least one dose). Therefore, we assume it will be difficult for active surveillance projects to detect those signals. Currently, vaccine safety monitoring projects using other databases were not able to detect statistically significant associations between their pre-specified outcomes and the COVID vaccine [12]. And the frequencies of reported AEs in those databases were consistent with results observed in clinical trials [31]. Therefore, we believe our system could benefit the vaccine surveillance community by magnifying signals for novel and rare AEs outside of the list of pre-specified outcomes routinely monitored by active surveillance projects.

The use of NLP is also critical in our system to magnify signals of rare AEs. For example, a vaccine recipient might document having a "fever," but a medical professional might diagnose the recipient with "pyrexia," which are different ways to refer to the same condition. Similarly, different expressions for "symptoms" (cough, muscle pain, congestion)

and a medical "diagnosis" (respiratory tract infection) adds additional complexity during data analysis.

We managed to control confounders and biases by 1) excluding patients already tested positive for COVID because symptoms from being positive for COVID might be a confounder for the detected AEs. 2) when calculating ROR, we used only influenza vaccine recipients as the control group instead of using recipients for all other vaccines because of the similarity shared between influenza and COVID. We assume that using the influenza vaccine as a control group may balance out symptoms as a sign of vaccine taking effect, which could potentially magnify signals for novel and rare AEs. We also attempted to investigate the strength of associative relationship revealed by ROR by providing a temporal (time-to-onset) information using our NLP system. The assumption is that shorter time-to-onset between AE and vaccination might be highly indicative of the association between the two events. This assumption has been adopted by recent studies [25, 27] when investigating the relationship between mRNA COVID vaccine and myocarditis.

However, our study only had limited ability to identify true causal relationships given the lack of confounding variables. In addition, since our population was based on people who developed AEs, ROR signals in Fig. 2 only represented a relative strength of a specific AE versus other AEs (especially routinely monitored AEs). This explained the relatively low RORs of J&J versus Pfizer/Moderna vaccines in cardiovascular AEs: J&J vaccines had more incidence of AEs in the routinely monitored category (e.g., fatigue, headache, nausea/vomit, pyrexia). Adding complementary data sources such as electronic health records (EHR) could allow the inclusion of people with and without AEs and enable more comprehensive assessment of vaccine AEs [32], which will be our future work. Nevertheless, cardiovascular AE signals were still significant for the COVID vaccine overall, thus worth additional verification using more robust vaccine databases.

V. CONCLUSION

In conclusion, we developed an NLP system to extract AE concepts and time-to-onset. The system was able to magnify signals for novel and rare AEs through 1) concept standardization and 2) ROR calculation with influenza vaccine recipients as the control group. By examining ROR as signals of AEs, we identified significant signals for cardiovascular and inflammatory AEs (e.g., appendicitis, pancreatitis, carditis). The signals were also verified by several recent case reports, which proved the utility of our proposed system for the early detection of novel and rare AEs. Despite the limitations in causality verification, our study provided an automatic system that enables timely identification of potential vaccine safety concerns, especially those outside of pre-specified outcomes routinely monitored.

ACKNOWLEDGMENT

The work is supported in part by NIH grants R01LM013337 and U01TR003528.

REFERENCES

- [1] T. T. Shimabukuro, M. Cole, and J. R. Su, "Reports of Anaphylaxis After Receipt of mRNA COVID-19 Vaccines in the US—December 14, 2020-January 18, 2021," *JAMA*, 2021.

- [2] R. T. Chen, S. C. Rastogi, J. R. Mullen, S. W. Hayes, S. L. Cochi, J. A. Donlon, and S. G. Wassilak, "The vaccine adverse event reporting system (VAERS)," *Vaccine*, vol. 12, no. 6, pp. 542-550, 1994.
- [3] T. T. Shimabukuro, "COVID-19 vaccine safety update," 2021.
- [4] J. Baggs, J. Gee, E. Lewis, G. Fowler, P. Benson, T. Lieu, A. Naleway, N. P. Klein, R. Baxter, and E. Belongia, "The Vaccine Safety Datalink: a model for monitoring immunization safety," *Pediatrics*, vol. 127, no. Supplement 1, pp. S45-S53, 2011.
- [5] R. L. Davis, M. Kolczak, E. Lewis, J. Nordin, M. Goodman, D. K. Shay, R. Platt, S. Black, H. Shinefield, and R. T. Chen, "Active surveillance of vaccine safety: a system to detect early signs of adverse events," *Epidemiology*, vol. 16, no. 3, pp. 336-341, 2005.
- [6] T. A. Lieu, M. Kulldorff, R. L. Davis, E. M. Lewis, E. Weintraub, K. Yih, R. Yin, J. S. Brown, R. Platt, and V. S. D. R. C. A. Team, "Real-time vaccine safety surveillance for the early detection of adverse events," *Medical care*, pp. S89-S95, 2007.
- [7] M. Weintraub, B. Fireman, J. Gee, K. Goddard, K. Hanson, B. Kieke, C. Lee, N. Lewis, D. McClure, and M. McNeil, "Rapid Cycle Analysis (RCA) to monitor the safety of COVID-19 vaccines in near real-time within the Vaccine Safety Datalink," 2021.
- [8] "V-safe protocol: version 2. Centers for Disease Control and Prevention.," Centers for Disease Control and Prevention.
- [9] J. Chapin-Bardales, J. Gee, and T. Myers, "Reactogenicity following receipt of mRNA-based COVID-19 vaccines," *Jama*, vol. 325, no. 21, pp. 2201-2202, 2021.
- [10] X. Wang, G. Hripcsak, M. Markatou, and C. Friedman, "Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study," *Journal of the American Medical Informatics Association*, vol. 16, no. 3, pp. 328-337, 2009.
- [11] T. T. Shimabukuro, M. Cole, and J. R. Su, "Reports of Anaphylaxis After Receipt of mRNA COVID-19 Vaccines in the US-December 14, 2020-January 18, 2021," *JAMA*, vol. 325, no. 11, pp. 1101-1102, Mar 16, 2021.
- [12] T. Shimabukuro. "COVID 19 vaccine safety update," March 31, 2021; <https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2021-02/28-03-01/05-covid-Shimabukuro.pdf>.
- [13] CDC. "Clinical immunization safety assessment (CISA) project," <https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/cisa/index.html>.
- [14] T. Botsis, T. Buttolph, M. D. Nguyen, S. Winiecki, E. J. Woo, and R. Ball, "Vaccine adverse event text mining system for extracting features from vaccine safety reports," *Journal of the American Medical Informatics Association*, vol. 19, no. 6, pp. 1011-1018, 2012.
- [15] J. Du, Y. Xiang, M. Sankaranarayananpillai, M. Zhang, J. Wang, Y. Si, H. A. Pham, H. Xu, Y. Chen, and C. Tao, "Extracting postmarketing adverse events from safety reports in the vaccine adverse event reporting system (VAERS) using deep learning," *Journal of the American Medical Informatics Association*, 2021.
- [16] K. J. Rothman, S. Lanes, and S. T. Sacks, "The reporting odds ratio and its advantages over the proportional reporting ratio," *Pharmacoepidemiology and drug safety*, vol. 13, no. 8, pp. 519-523, 2004.
- [17] "COVID-19 and the Flu," 2021; <https://asm.org/Articles/2020/July/COVID-19-and-the-Flu>.
- [18] Y. Li, B. Liu, J. Cui, Z. Wang, Y. Shen, Y. Xu, K. Yao, and Y. Guan, "Similarities and evolutionary relationships of COVID-19 and related viruses," *arXiv preprint arXiv:2003.05580*, 2020.
- [19] A. R. Aronson, "Metamap: Mapping text to the umls metathesaurus," Bethesda, MD: NLM, NIH, DHHS, pp. 1-26, 2006.
- [20] H. Liu, S. J. Bielinski, S. Sohn, S. Murphy, K. B. Waghlikar, S. R. Jonnalagadda, K. Ravikumar, S. T. Wu, I. J. Kullo, and C. G. Chute, "An information extraction framework for cohort identification using electronic health records," *AMIA Summits on Translational Science Proceedings*, vol. 2013, pp. 149, 2013.
- [21] A. Dionne, F. Sperotto, S. Chamberlain, A. L. Baker, A. J. Powell, A. Prakash, D. A. Castellanos, S. F. Saleeb, S. D. de Ferranti, and J. W. Newburger, "Association of Myocarditis With BNT162b2 Messenger RNA COVID-19 Vaccine in a Case Series of Children," *JAMA cardiology*, 2021.
- [22] K. Watkins, G. Griffin, K. Septaric, and E. L. Simon, "Myocarditis after BNT162b2 vaccination in a healthy male," *The American Journal of Emergency Medicine*, 2021.
- [23] A. Isaak, A. Feisst, and J. A. Luetkens, "Myocarditis Following COVID-19 Vaccination," *Radiology*, pp. 211766, 2021.
- [24] G. A. Diaz, G. T. Parsons, S. K. Gering, A. R. Meier, I. V. Hutchinson, and A. Robicsek, "Myocarditis and pericarditis after vaccination for COVID-19," *JAMA*, 2021.
- [25] J. Montgomery, M. Ryan, R. Engler, D. Hoffman, B. McClenathan, L. Collins, D. Loran, D. Hrnecir, K. Herring, and M. Platzer, "Myocarditis following immunization with mRNA COVID-19 vaccines in members of the US military," *JAMA cardiology*, 2021.
- [26] J. B. García, P. P. Ortega, A. B. Fernández, A. C. León, L. R. Burgos, and E. C. Dorta, "Acute myocarditis after administration of the BNT162b2 vaccine against COVID-19," *Revista Espanola de Cardiologia*, 2021.
- [27] H. W. Kim, E. R. Jenista, D. C. Wendell, C. F. Azevedo, M. J. Campbell, S. N. Darty, M. A. Parker, and R. J. Kim, "Patients with acute myocarditis following mRNA COVID-19 vaccination," *JAMA cardiology*, 2021.
- [28] B. Bozkurt, I. Kamat, and P. J. Hotez, "Myocarditis with COVID-19 mRNA Vaccines," *Circulation*, vol. 144, no. 6, pp. 471-484, 2021.
- [29] H. Merchant, "CoViD Vaccines and thrombotic events: Possibility of mRNA translation and spike protein synthesis by platelets?," *The BMJ*, vol. 372, pp. n699/tr-6, 2021.
- [30] J. W. Gargano, M. Wallace, S. C. Hadler, G. Langley, J. R. Su, M. E. Oster, K. R. Broder, J. Gee, E. Weintraub, and T. Shimabukuro, "Use of mRNA COVID-19 vaccine after reports of myocarditis among vaccine recipients: update from the Advisory Committee on Immunization Practices—United States, June 2021," *Morbidity and Mortality Weekly Report*, vol. 70, no. 27, pp. 977, 2021.
- [31] J. Chapin-Bardales, J. Gee, and T. Myers, "Reactogenicity Following Receipt of mRNA-Based COVID-19 Vaccines," *JAMA*, 2021.
- [32] Y. Luo, W. K. Thompson, T. M. Herr, Z. Zeng, M. A. Berendsen, S. R. Jonnalagadda, M. B. Carson, and J. Starren, "Natural language processing for EHR-based pharmacovigilance: a structured review," *Drug safety*, vol. 40, no. 11, pp. 1075-1089, 2017.