Clustering Adverse Events of COVID-19 Vaccines across the United States

Ahmed Askar[1,2], MPH, PMP and Andreas Züfle[1], PhD
George Mason University[1] and U.S. FDA[2]

ABSTRACT

We study the similarity of adverse effects of COVID-19 vaccines across different states in the United States. We use data of ~300,000 COVID-19 vaccine adverse event reports obtained from the Vaccine Adverse Event Reporting System (VAERS). We extract latent topics from the reported adverse events using a topic modeling approach based on Latent Dirichlet allocation (LDA). This approach allows us to represent each U.S state as a low-dimensional distribution over topics. Using Moran's index of spatial autocorrelation we show that some of the topics of adverse events exhibit significant spatial autocorrelation, indicating that there exist spatial clusters of nearby states that exhibit similar adverse events. Using Anselin's local indicator of spatial association we discover and report these clusters. Our results show that adverse events of COVID-19 vaccines vary across states which justifies further research to understand the underlying causality to better understand adverse effects and to reduce vaccine hesitancy.

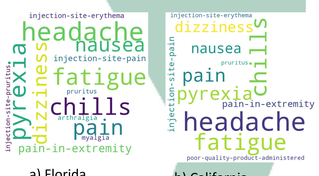
INTRODUCTION

By June 12th, 2021, more than 2.3 billion doses of various brands of COVID-19 vaccines had been administered world-wide with more than 300 million doses administered in the United States[1]. The U.S. Centers for Disease Control and Prevention (CDC) has stated that all U.S. authorized vaccines are safe and efficient[2].While generally safe, the COVID-19 vaccines have adverse effects, including common side effects such as injection site pain and fever, but also including rare adverse effects that can be more severe. In our prior work in [3] we performed a spatio-temporal study on the adverse events of blood thinning drugs and their spatial auto-correlation.

RESEARCH QUESTION/HYPOTHESIS

- To identify adverse events associated with FDA approved drugs/vaccines and identify any spatial patterns.

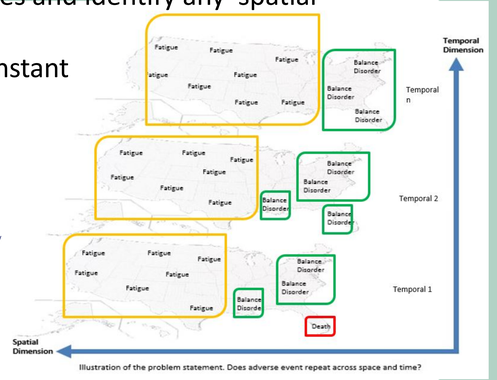
❖Are adverse events constant for each spatial region.



a) Florida

b) California

Fig 1) COVID-19 Adverse Effect Clouds per Region



METHODOLOGY

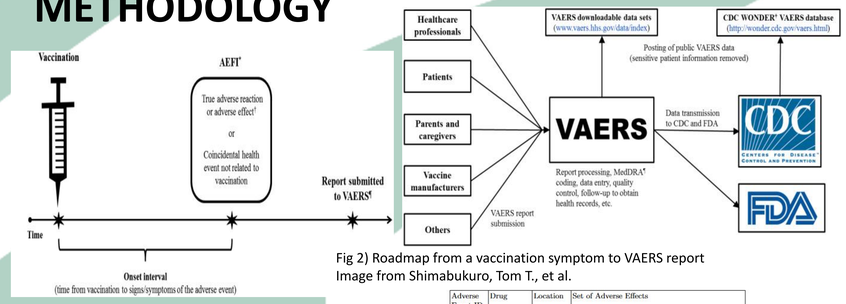


Fig 2) Roadmap from a vaccination symptom to VAERS report
Image from Shimabukuro, Tom T., et al.

Adverse Event ID	Drug	Location	Set of Adverse Effects
1139067	Moderna	MD	Dizziness, Injection site pruritus, Injection site rash, Somnolence
1004857	Moderna	PA	Nausea, Palpitations, Presyncope, Pyrexia, Tremor
1115746	Moderna	NY	Chills,Headache,Nausea,Pain,Pain in extremity
1148711	Moderna	CA	Axillary pain, Fatigue, Headache, Nausea, Pain in extremity
1240185	Pfizer	IN	Fatigue,Headache,Pain,Pyrexia
1120846	Pfizer	UT	Nausea,Pain in extremity, Sleep disorder, Tinnitus, Vertigo
1104541	Pfizer	GA	Injection site reaction, Rash pruritic
1138693	Pfizer	WI	Eye pruritus, Lip swelling, Nasal pruritus, Swelling face, Urticaria
1200860	Janssen	TX	Headache
1114482	Janssen	MI	Chills, Hyperhidrosis, Pyrexia
1244933	Janssen	IL	Heart rate, Heart rate increased, Pain, Poor quality sleep, Pyrexia
1202067	Janssen	RI	Chills, Injection site erythema, Menstruation irregular, Pyrexia

Table 1: Sample records of Adverse Event Report Database. Each Line is an Adverse Event.

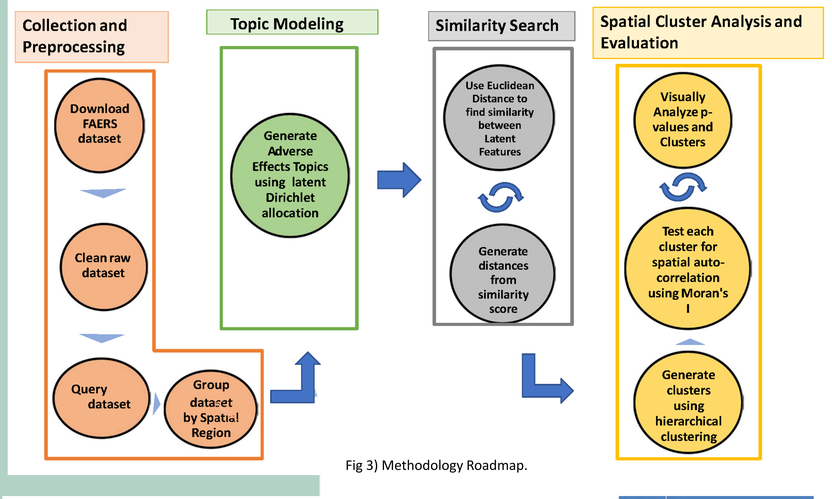
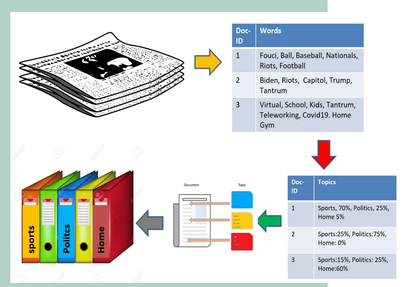


Fig 3) Methodology Roadmap.



Doc-ID	Words
1	Fouci, Ball, Baseball, Nationals, Riots, Football
2	Biden, Riots, Capitol, Trump, Tantrum
3	Virtual, School, Kids, Tantrum, Teleworking, Covid19. Home Gym

Doc-ID	Topics
1	Sports, 70%, Politics, 25%, Home 5%
2	Sports:25%, Politics:75%, Home: 0%
3	Sports:15%, Politics: 25%, Home:60%

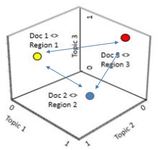
Fig 4) Latent Dirichlet Allocation

Region 1 vs Region n

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

P-Q = two points in Euclidean n-space
q_i, p_i = Euclidean vectors, starting from the origin of the space (initial point)
n = n-space

Reg-ID	Topics
1	Fall: 0.7, Bleeding:0.25, Stroke: 0.05...n topics
2	Stroke:0.25, Bleeding:0.75 Fall: 0...n topics
3	Bleeding:0.15 Fall: 0.25, Stroke:0.6...n topics



	Reg-1	Reg-2	Reg-3
Region1	0	0.67	0.77
Region 2	0.67	0	0.78
Region3	0.77	0.78	0

Fig 5) Similarity distance between topics

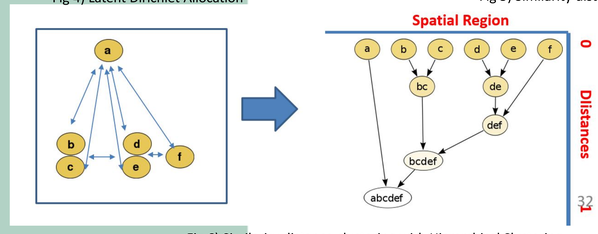


Fig 6) Similarity distance clustering with Hierarchical Clustering

RESULTS

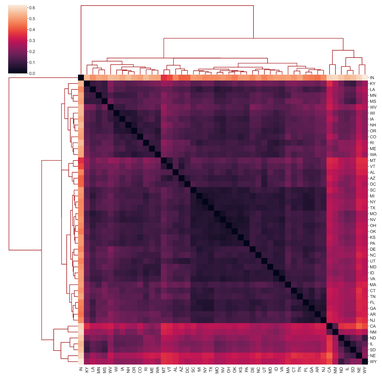
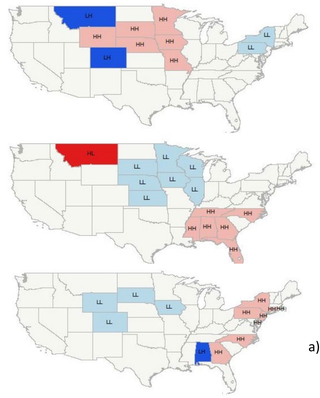


Fig 7) Pair-wise similarity matrix of latent topics of COVID-19 vaccine adverse events of counties in the United States.

Pattern	p-value	Moran's Index	z-score	Topic ID
Clustered	0.0006	0.2756	3.4512	1
Random	0.6214	-0.0635	-0.4938	2
Clustered	0.0966	0.1216	1.6616	3
Random	0.6643	-0.0464	-0.4340	4
Random	0.2054	0.0920	1.2662	5
Random	0.6867	0.0109	0.4033	6
Dispersed	0.0754	-0.1785	-1.7782	7
Clustered	0.0071	0.2149	2.6938	8
Random	0.1988	0.0875	1.2850	9
Clustered	0.0002	0.3163	3.7895	10

Table 3: Moran's I measure of global spatial autocorrelation for each of the $K = 10$ topics of COVID-19 adverse events.



a) Topic # 1

a) Topic # 8

a) Topic # 10

Fig 8) Local Indicator of Spatial Autocorrelation (LISA). Light red areas correspond to high-high clusters. Light blue areas are low-low clusters. Dark red and dark blue areas corresponds to high-low and low-high outliers.

DISCUSSION/CONCLUSIONS

In this work, we tackled the problem of measuring (dis-)similarity between adverse events of COVID-19 vaccines observed in different regions. Our measure leverages a topic modeling approach using LDA to map each adverse event from a (textual) set of adverse effects to a latent topic distribution. Using a database of ~300,000 adverse event reports of COVID-19 vaccines in the United States, investigate the underlying topics exhibit any spatial autocorrelation to understand if different places exhibit different adverse events. Our results show that some of the latent topics of COVID-19 adverse events show significant positive spatial autocorrelation. Our local analysis of spatial autocorrelation show that certain topics of adverse events have increased (or decreased) likelihood in different parts of the United States. We hope that teams of medical experts may find this result to investigate the underlying causality. Reasons could be due to vaccine quality issues, storage and cooling issues, or simply due to different brands of vaccines.

ACKNOWLEDGEMENTS

This research was prepared or accomplished by Ahmed Askar in his personal capacity. The opinions expressed in this article are the author's own and do not reflect the view of the U.S Food and Drug Administration, the Department of Health and Human Services, or the United States government. This research received no external funding.

REFERENCES:

1) Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to trackCOVID-19 in real time. (https://coronavirus.jhu.edu/map.html). The Lancet in-fectious diseases20(5), 533–534 (2020)
2) Centers for Disease Control and Prevention: Different COVID-19 Vaccines.(https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html)
3) Askar, Ahmed, and Andreas Zuefle. "Clustering of Adverse Events of Post-Market Approved Drugs." *17th International Symposium on Spatial and Temporal Databases*. 2021.