Improving Pattern Discovery in the FDA Adverse Event Reporting System (FAERS) with Network Analysis (NA)

Raechel Davis[1,2], Oanh Dang[1], Suranjan De[1], Robert Ball[1]

1. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, U.S. Food and Drug Administration
2. Oak Ridge Institute for Science and Education

Funding: RD is supported by appointment to the Research Participation Program at the U.S. Food and Drug Administration Center for Drug Evaluation and Research, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

Acknowledgments: We thank the RAPID, FAERS, and InfoViP teams for their support.

Abstract

Background: In drug-safety monitoring systems, adverse events (AEs) associated with the use of medical products often consist of complex patterns of clinical events. NA was used for pattern recognition and characterizing the Vaccine Adverse Event Reporting System (VAERS), but limited applications to FAERS left its network description incomplete. In this analysis, the network properties of FAERS were characterized and leveraged to facilitate pattern discovery.

Methods: FAERS subsets were analyzed with the drugs/AEs as the nodes and their interconnections as the edges. FAERS was evaluated to determine if it has scale-free network characteristics, with central nodes and similar structures. Metrics that evaluate network connectivity and edge weighting algorithms for community detection were applied. Network-based signal detection metrics were used to identify potential adverse drug events. The analyses were conducted in R using the package "igraph" on data from a copy of the FAERS database in the FDA/CDER RAPID platform.

Results: Serious reports from 2016-2023 (2,062,099) were represented as a network of 20,965 nodes (16,847 AEs and 4,116 drugs) with more than four million interconnections. FAERS subnetworks were determined to have scale-free characteristics and small-world phenomenon with heavy tailed degree distributions, high local clustering, and low diameters. Community detection techniques and network-based signal detection metrics identified clustering patterns representative of AEs in FAERS subsets.

Conclusion: To the best of our knowledge, this is the first systematic application of NA to FAERS that provides insight into the overall network properties of the database and evaluates AEs across therapeutic areas and report types.

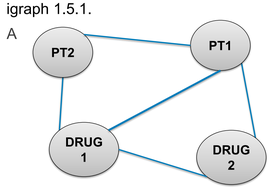
Introduction

- Networks are structures representing a group of elements (nodes) and the connections (edges) between them. Statistical and structural network properties can provide insights into relationships among nodes.
- The first application of NA to VAERS revealed scale-free network characteristics and explored vaccines and AEs involved in known safety signals.
- FAERS contains reports of suspected AEs, medication errors, and product quality issues related to drugs and biologic products.
- NA applied to FAERS may reveal statistical properties and multidimensional relationships among drugs and AEs to facilitate exploration of patterns of potential clinical interest.

Materials and Methods

- Data used in this study is acquired from a copy of the Information Visualization Platform (InfoViP) FAERS database. All data manipulation was performed in the FDA/CDER RAPID platform using R 4.2.3.
- Serious, domestic and deduplicated reports from 2016-2023 were extracted. The key variables used include: FAERS unique identifier, MedDRA Preferred Terms (PTs), and product active ingredients (PAIs). Drug PAIs were primarily mapped to their active moieties with some broader groupings of similar products.
- To investigate different report types, a subset of only spontaneous reports was obtained from the extracted reports. To explore network-based signal detection for a known adverse event, a subset of reports containing a PT for "rhabdomyolysis" were obtained.
- FAERS subsets were processed individually so that each report was represented as a vector consisting of drugs and PTs which were then decomposed into all possible pairs.
- The number of reports containing the same pair was the weight for each element, known as the report-based weight. For signal detection, triangle weighting was also implemented where the weight is the number of edges between two nodes that are part of a triangle.
- These weighted pairs were presented in a three-column matrice known as an edge list, used to generate the network (Figure 1). Network generation and evaluation was performed using igraph 1.5.1.

A



B

Pair		RW	TW
DRUG1	DRUG2	1	1
PT1	DRUG1	2	2
PT2	DRUG1	1	1
DRUG2	PT2	1	0
PT2	PT1	1	1
PT1	DRUG2	1	1

Figure 1. Example networks built by different edge weighting algorithms for the following reports: Report 1 = [DRUG1 DRUG2 PT1 PT2] and Report 2 = [DRUG1 PT1]. (A) Networks visualized from the reports. (B) Corresponding edge lists. RW, report-weight; TW, triangle-weight

- Network characteristics were calculated: degree (number of connections), strength (weighted degree), average path length (average distance between any two nodes), diameter (longest shortest path), global transitivity (count of triangles), and average local clustering coefficient (proportion of all possible connections). The degree distribution was visualized.
- The rhabdomyolysis network centrality was evaluated, and the network was visualized using the different edge weights.

Results and Discussion

- FAERS network characteristics (Table 1) are directly compared to the network characteristics of VAERS *[Ball R., Botsis T., 2011].*
- FAERS is a dense network, with any given node connected to all other nodes through an average of approximately two other nodes and a maximum of five nodes.

Table 1. General characteristics of the networks

Network	$N_{reports}$	N_{nodes}	N_{edges}	N_{PTs}	N_{drugs}	$\bar{s}$	$\bar{k}$	$\bar{l}$	∅	Transitivity	$\overline{CC}$
Full FAERS	2,063,286	20,965	4,052,196	16,847	4,116	3208.9	386.6	2.23	5	0.286	0.766
Spon	1,340,386	17,945	2,750,845	14,573	3,370	2,152.8	306.6	2.27	5	0.275	0.752
Rhabdo	2,466	2,400	67,075	1,670	730	104.9	55.9	1.98	2	0.223	0.765
VAERS	319,424	6,428	67,075	6,354	74	3,600.7	-	2.22	6	-	0.771

$N_{reports}$, number of reports; N_{nodes}, number of nodes; N_{edges}, number of edges; N_{PTs} number of PTs; N_{drugs} number of drugs; $\bar{s}$, average strength of nodes; $\bar{k}$, average degree of nodes; $\bar{l}$, average path length; ∅, network diameter; global transitivity; $\overline{CC}$, average clustering coefficient; Spon, Spontaneous; Rhabdo, Rhabdomyolysis; PTs, Standardized Medical Dictionary for Regulatory Activities Preferred Terms.

- FAERS subnetworks have very similar properties and were determined to have small-world phenomenon with heavy tailed degree distributions, high local clustering (~0.8), and low diameters ranging from 2-5 [*Gupta R., et al., 2016].*
- The average local clustering coefficients of FAERS and VAERS are similar suggesting similar network density and community presence.
- Many elements in FAERS (~80%) and VAERS (~99%) are PTs. The PT-rich nature of the networks are likely because many PTs are reported together, as adverse events often appear as multiple clinical symptoms.
- FAERS has characteristics of a scale-free network where most nodes have few connections and certain drugs and AEs may behave as "hubs" (Figure 2).

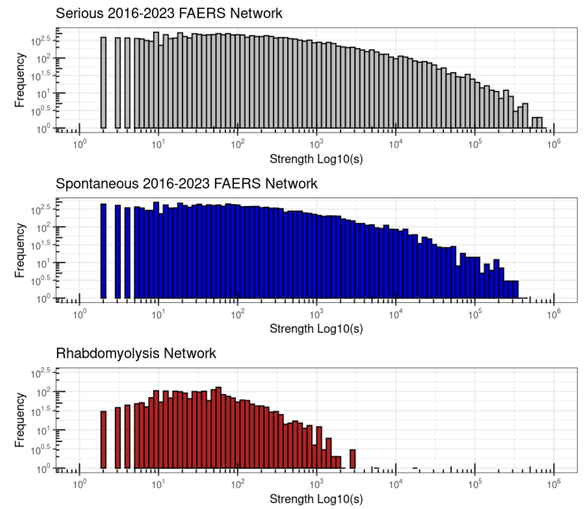


Figure 2. Degree distributions of the networks of counts of reports by degree on a log-log scale.

- Metrics that identify central nodes in the network reveal PTs that are part of the Standardized MedDRA Query (SMQ) for rhabdomyolysis and statin drug products known to be associated with the condition (Figure 3).

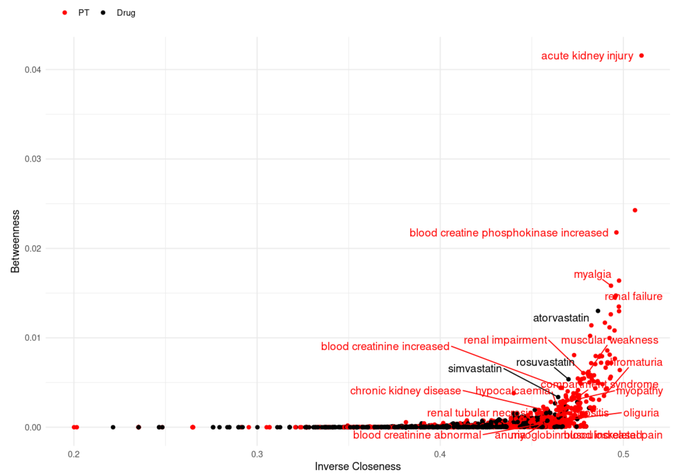


Figure 3. Node centrality metrics as measures of connectedness for the rhabdomyolysis subnetwork of FAERS. The red points are PTs, and the black points are drug products. Certain nodes of interest are labelled. Betweenness centrality measures the extent to which each node falls between other nodes and inverse-closeness centrality measures the mean distance from a given node to other nodes. The "rhabdomyolysis" node was a highly central outlier due to its presence in reports for inclusion into the subnetwork and was removed from the figure.

[Ball, R. and T. Botsis, Can Network Analysis Improve Pattern Recognition Among Adverse Events Following Immunization Reported to VAERS? Clinical Pharmacology & Therapeutics, 2011. 90(2): p. 271-278.]

[Gupta, R., T. Roughgarden, and C. Seshadhri, Decompositions of Triangle-Dense Graphs. Siam Journal on Computing, 2016. 45(2): p. 197-215.]

- Community detection techniques identified clustering patterns representative of AEs in FAERS subsets (Figure 4).

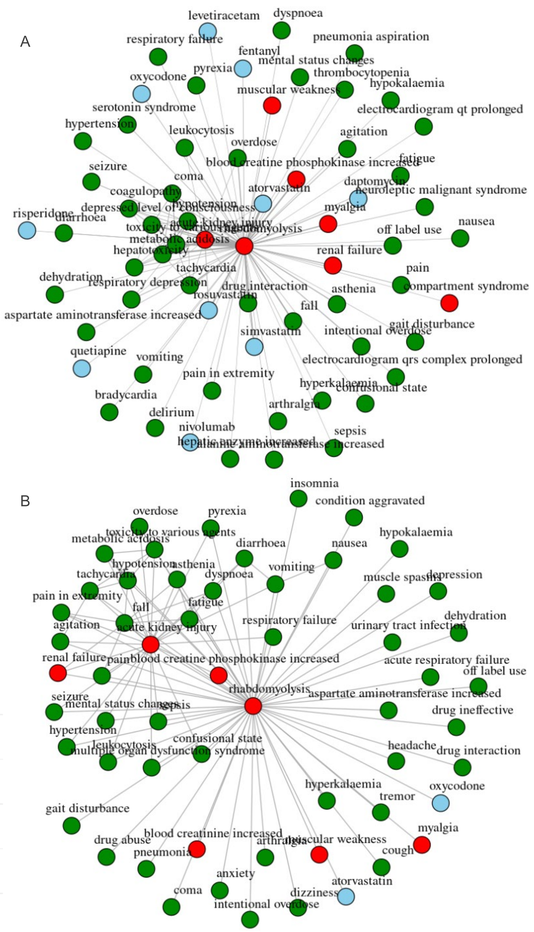


Figure 4. Network visualizations of the top 100 weighted edges demonstrating a pattern of rhabdomyolysis related PTs. Green circles designate PTs, red circles reveal PTs that are part of the SMQ for rhabdomyolysis, and blue circles designate drug products. (A) Report-based edge weighted network. (B) Triangle edge weighted network

Conclusions

- For the first time to our knowledge, large subsets of FAERS are shown to have characteristics of a scale-free network where some drugs and PTs are highly connected and central to the network.
- Highly weighted connections between drugs and PTs in the rhabdomyolysis subnetwork represent relationships of known safety concern and clinical significance, particularly the presence of statins.
- NA offers a complementary approach to support pharmacovigilance by exploring and visualizing the interrelationships present in FAERS data at the level of the drug(s), AE(s), or the combination thereof.