Drug Safety (2025) 48:513–526
https://doi.org/10.1007/s40264-025-01515-y

ORIGINAL RESEARCH ARTICLE

Network Analysis and Machine Learning for Signal Detection and Prioritization Using Electronic Healthcare Records and Administrative Databases: A Proof of Concept in Drug-Induced Acute Myocardial Infarction

Maria Antonietta Barbieri[1,2] · Andrea Abate[1,2] · Olivér M. Balogh[3,4] · Mátyás Pétervári[3,4,5] · Péter Ferdinandy[3,4,6] · Bence Ágg[3,4,6] · Vera Battini[7] · Marianna Cocco[7] · Andrea Rossi[8,9] · Carla Carnovale[7] · Manuela Casula[8,9] · Edoardo Spina[1] · Maurizio Sessa[2]

Accepted: 16 January 2025 / Published online: 7 February 2025

Abstract

Background Safety signals for potential drug-induced adverse events (AEs) typically emerge from multiple data sources, primarily spontaneous reporting systems, despite known limitations. Increasingly, real-world data from sources such as electronic health records (EHRs) and administrative databases are leveraged for signal detection. Although network analysis has shown promise in mapping relationships between clinical attributes for signal detection in spontaneous reporting system databases, its application in real-world data from EHRs and administrative databases remains limited.

Objective This study aimed to evaluate the performance of network analysis in detecting safety signals within Italian administrative databases, using drug-induced acute myocardial infarction (AMI) as a proof of concept.

Methods We employed a case–crossover design to explore the association between drug exposure and AMI using the Healthcare Administrative Database of Mantova, Italy, from 2014 to 2018. Patients with their first AMI hospitalization were identified after a 365-day washout period to exclude prior hospitalizations. We constructed a network to analyse the relationships between prescribed drugs and diagnoses, represented as nodes, with undirected edges illustrating their interactions. For each patient with AMI, we identified all diagnoses and drugs recorded or redeemed within 365 days of the first AMI episode and generated various drug–diagnosis, drug–drug, and diagnosis–diagnosis pairs. We calculated the frequency of these pairs, and three types of edge weights quantified the strength of connections. We identified outlier drug–AMI pairs using a predictive score (F) based on frequency (C) and full edge weights (W_F), with validation for known AMI associations. We prioritized signals using the F score, C of AMI, and W_F, analysed through k-means clustering to identify patterns in the data.

Results From 2014 to 2018, a total of 3918 patients had an AMI, with 4686 AMI diagnoses. Of those, 2866 had prescriptions in the previous year, totalling 498,591 prescriptions. A network analysis identified 2968 unique nodes, revealing 529,935 diagnosis–diagnosis connections, 235,380 drug–diagnosis connections, and 102,831 drug–drug connections. The median number of connections (C) was 404 ($Q1$–$Q3$: 194–671) for drug nodes and 380 ($Q1$–$Q3$: 216–664) for diagnosis nodes. The median W_F was 11.8 ($Q1$–$Q3$: 9–14), and the median F score across pairs was 0.1 ($Q1$–$Q3$: 0.1–0.3). A total of 249 potential safety signals were detected, with 63.4% aligning with known AEs. Among the remaining signals, 80 were prioritized, and five emerged as the highest priority: terazosin, tamsulosin, allopurinol, esomeprazole, and omeprazole.

Conclusions Overall, our novel method demonstrates that network analysis is a valuable tool for signal detection and prioritization in drug-induced AEs based on EHRs and administrative databases.

Extended author information available on the last page of the article

Key Points

Spontaneous reporting systems databases are the cornerstone for signal detection, but their limitations have led to the exploration of real-world data, such as claims data and electronic health records, for enhancing signal detection capabilities.

The application of network analysis in signal detection, especially using electronic health records and administrative databases, is still novel and largely unexplored.

We constructed a network of 2968 nodes representing drugs and diagnoses, which identified over 800,000 connections. A total of 249 potential drug–acute myocardial infarction signals were detected, with 63.4% aligning with a known adverse event.

1 Introduction

Safety signals refer to information on a new or known adverse event (AE) that is potentially caused by a medicine or a vaccine and that warrants further investigation. Signals are generated from several data sources, including spontaneous reports, clinical studies, and the scientific literature. Spontaneous reporting systems (SRS) databases continue to be the cornerstone data source in signal detection, despite their intrinsic limitations [1–4]. Other complementary data sources are now widely explored for use in signal detection, including real-world data (RWD), encompassing claims data, electronic health records (EHRs), and administrative databases [5–7]. The Sentinel System in the USA and DARWIN-EU in the European Union are largely distributed database networks that also actively analyse RWD for signal detection and validation, facilitating timely interventions by employing common data models and analytical tools [8–12]. These distributed database networks use a variety of methods and analytical approaches. No single method/approach has demonstrated superior performance for signal detection across all drugs/vaccines and outcomes when using EHRs and administrative databases, highlighting the need for a wide variety of methods/approaches that can be used in different scenarios [7, 13–16].

Network analysis, a mathematical framework used to study the relationships (i.e., edges) between individual attributes (i.e., nodes) [17, 18] has been recently explored as a complementary method for signal detection in SRS databases with a proof of concept in drug-induced AEs, including acute myocardial infarction (AMI) [19–27]. Network analysis has been widely explored in the studies of disease progression and drug prescription pattern using EHRs and claims data [19, 20, 27–30]. However, to the best of our knowledge, no studies have been conducted to adapt network analysis for the purpose of signal detection through EHRs and administrative databases. Network analyses for signal detection in EHRs offers promise as an alternative to SRS by leveraging comprehensive RWD, useful to identify complex patterns and interactions between clinical events, medications, and patient characteristics [7, 31, 32]. Therefore, we investigated the use and the performance of network analysis as a method for signal detection in RWD, focusing on Italian EHRs and administrative databases using drug-induced AMI as a proof of concept.

2 Methods

2.1 Data Source

The data used in this study were sourced from the Healthcare Administrative Database of the Lombardy Region, which encompasses records for all residents within the region, constituting nearly 10 million individuals, approximately 16% of the Italian population [33]. Due to limited computational resources, the study is based on data from 2014 to 2018 related to the Healthcare Administrative Database from Mantova, Italy. All data are from both public and private accredited healthcare providers, partly or entirely funded by the Italian National Health Service (NHS). The database incorporates several components: (a) the archive of NHS beneficiaries, encompassing essentially the entire resident population and containing their demographic and administrative characteristics; (b) the hospital discharge database, documenting all hospital admissions across public and private hospitals in Lombardy with primary and secondary discharge diagnoses; and (c) the outpatient prescription drug database, containing information on NHS-reimbursable medication prescriptions. Integration of diverse pieces of information pertaining to NHS beneficiaries, stored across these repositories, is facilitated through a unique personal identification code. Adhering to prevailing Italian legislation on patient confidentiality, the unique identification code undergoes automatic conversion to an anonymous code, with reversal of this process pre-empted by deletion of the conversion table. Further elucidation regarding the structure of this database is available elsewhere [34].

All diagnoses are classified in accordance with the International Statistical Classification of Diseases and Related Health Problems, 9th revision, Clinical Modification (ICD-9-CM) [35], and drug prescriptions are categorized according to the fifth level of the Anatomical Therapeutic Chemical (ATC) classification system codes [36]. Data obtained from this database have been validated previously [34, 37–40].

2.2 Study Design

This study employed a case–crossover design. Individuals who experienced their first hospitalization for AMI during the study period were identified, without age limits. The study period was selected for data availability. A 365-day washout period was applied to ensure no prior AMI hospitalizations in the previous year, defining this as the first AMI. As the 1-year look-back period is necessary, enrolment started in 2015. Both primary and secondary diagnoses documented during hospital admissions, using ICD-9 CM codes beginning with 410, were examined to identify AMI. The positive predictive value of code 410, specifically when used as the primary diagnosis, was reported to be 86% [41]. Within a 365-day observational window leading up to the AMI event, individuals served as their own controls (Fig. 1) [42].

2.3 Network Analysis

We constructed a network to map the relationships between various elements, with prescribed drugs and diagnoses representing the nodes. The connections between these nodes were depicted as undirected edges, illustrating the interactions between the drugs and diagnoses. Among all patients with AMI, we identified all other diagnoses and drugs recorded/redeemed within 365 days of their first AMI episode within the study period. Next, we created all possible combinations of drugs and diagnoses (i.e., drug–diagnosis pairs) that could be identified within this observational window of 365 days. We then calculated the frequency of each pair (C), representing the number of patients who had both the specific drug and diagnosis as a co-occurrence. This same pairing approach was applied to identify all combinations of dispensed drugs (i.e., drug–drug pairs) and diagnosis codes (i.e., diagnosis–diagnosis pairs) (Fig. 2). To quantify the strength of the connection between nodes (size of the edge in Fig. 2), three types of edge weights were computed for each pair (i.e., drug–drug, drug–diagnosis, and diagnosis–diagnosis):

- *Drug edge weight (W_{Dr}):* To calculate W_{Dr}, for each pair (i.e., drug–drug, drug–diagnosis, and diagnosis–diagnosis), we identified all related patients with the given pair; then, for each patient, we determined how many other drugs were co-dispensed within the 365 observational window. From this, we created a distribution and calculated the 75th percentile (quartile 3 [$Q3$]) with a continuity correction of 0.5 to obtain the final W_{Dr} value.
- *Diagnosis edge weight (W_{Di}):* To calculate W_{Di}, for each pair (i.e., drug–drug, drug–diagnosis, and diagnosis–diagnosis), we determined how many other diagnoses were co-recorded within the 365 observational window and calculated the 75th percentile ($Q3$) of this distribution with a continuity correction of 0.5 to obtain the final W_{Di} value.
- *Full edge weight (W_F):* the W_{Dr} and W_{Di} values were summed to derive the overall W_F for each pair (i.e., drug–drug, drug–diagnosis, and diagnosis–diagnosis). This gave the number of other diagnoses and drugs with a certain edge. There are epidemiological reasons to compute the W_F scores. Specifically, W_F considers at the same time the number of drugs co-dispensed alongside the prevalence of co-recorded diagnoses. Drugs with an edge with AMI that have fewer co-dispensed medications and co-recorded diagnoses (and, therefore, a lower W_F) have less potential for alternative 'causes' for their association with AMI and therefore a stronger epidemiological association. Conversely, the presence of multiple co-dispensed drugs, as well as additional co-recorded diagnoses, suggest potential alternative 'causes' for AMI and therefore indicate a weakened association.

A graphical representation of the datasets generated from these calculations is provided in Fig. S1 in the electronic supplementary material.

2.4 Signal Detection

The signal detection was conducted through outlier detection. Outlier detection was performed to identify drug–AMI pairs with unusually high frequencies relative to their W_F scores. Initially, the final predictive score (F) was computed for each drug–diagnosis pair, representing the ratio of frequency (C) to the respective W_F value through the following equation:

$$F = \frac{C}{W_F}.$$

Outliers of the F score were identified using quartiles 1 and 3 ($Q1$–$Q3$). Pairs with F that exceeded 1.5 times the $Q3$ were classified as outliers. Detected outliers were further analysed to ascertain potential signals of interest. A density plot was generated to visualize the distribution of outlier scores, aiding in the identification of significant signals within the dataset.

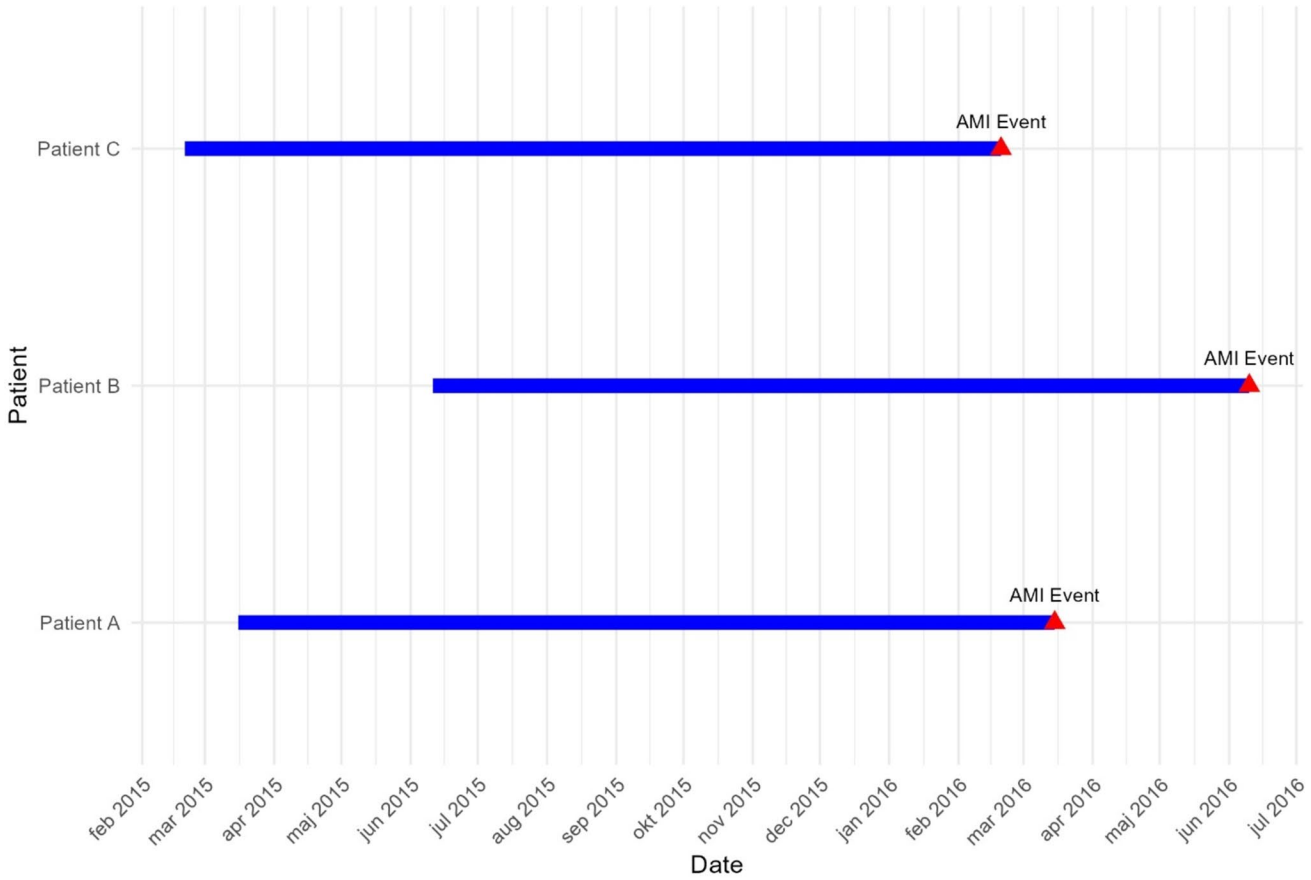

Fig. 1 Case–crossover design: Each patient's 1-year timeline (blue line) depicts the period leading up to the diagnosis of acute myocardial infarction (AMI) (red triangle)

To evaluate the performance of our signal detection method, we used the SIDER database [43]. We considered in SIDER all drugs that listed in their summary of product characteristics (SmPCs) the Medical Dictionary for Regulatory Activities preferred term 'myocardial infarction' to be true positives. Since our data sources were pharmacy databases containing only drugs reimbursed by the Italian NHS, we excluded drugs not reimbursed or dispensed at community pharmacies from the list of drugs extracted from SIDER. These excluded drugs dispensed only in hospitals and those whose cost is fully borne by the patient (both with and without medical prescription) [44]. Additionally, considering that SIDER is based on the US Food and Drug Administration (FDA) prescribing information, we excluded from the list of drugs extracted from SIDER those not commercialized in Italy during the study period. To assess the performance of our approach, we calculated the percentage of drugs identified by our method relative to the filtered list of drugs obtained from SIDER that listed 'myocardial infarction' (denominator) in the SmPC. Following identification as potential safety signals through an initial screening process, the drugs undergoing validation were categorized based on their therapeutic use from the ATC level. These categories included antibacterials for systemic use, analgesics, anti-inflammatory and antirheumatic products, drugs used in diabetes, antihypertensives, psychoanaleptics, drugs for gastrointestinal and respiratory conditions, and vitamins and supplements. Broader categories captured antipsychotics for treating mental illness, anticonvulsants for seizure control, ophthalmic agents for eye care, bone metabolism agents, hormones, and a miscellaneous category for less common therapeutic uses.

2.5 Signal Validation

We conducted signal validation to assess whether AMI was a known or unknown AE associated with each drug. This validation process involved consulting the European Medicines Agency SmPC [45] or Italian Medicines Agency SmPC if the drug is not authorized by centralized procedure, and the US FDA full prescribing information [46] to check whether the preferred term 'myocardial infarction' was listed.

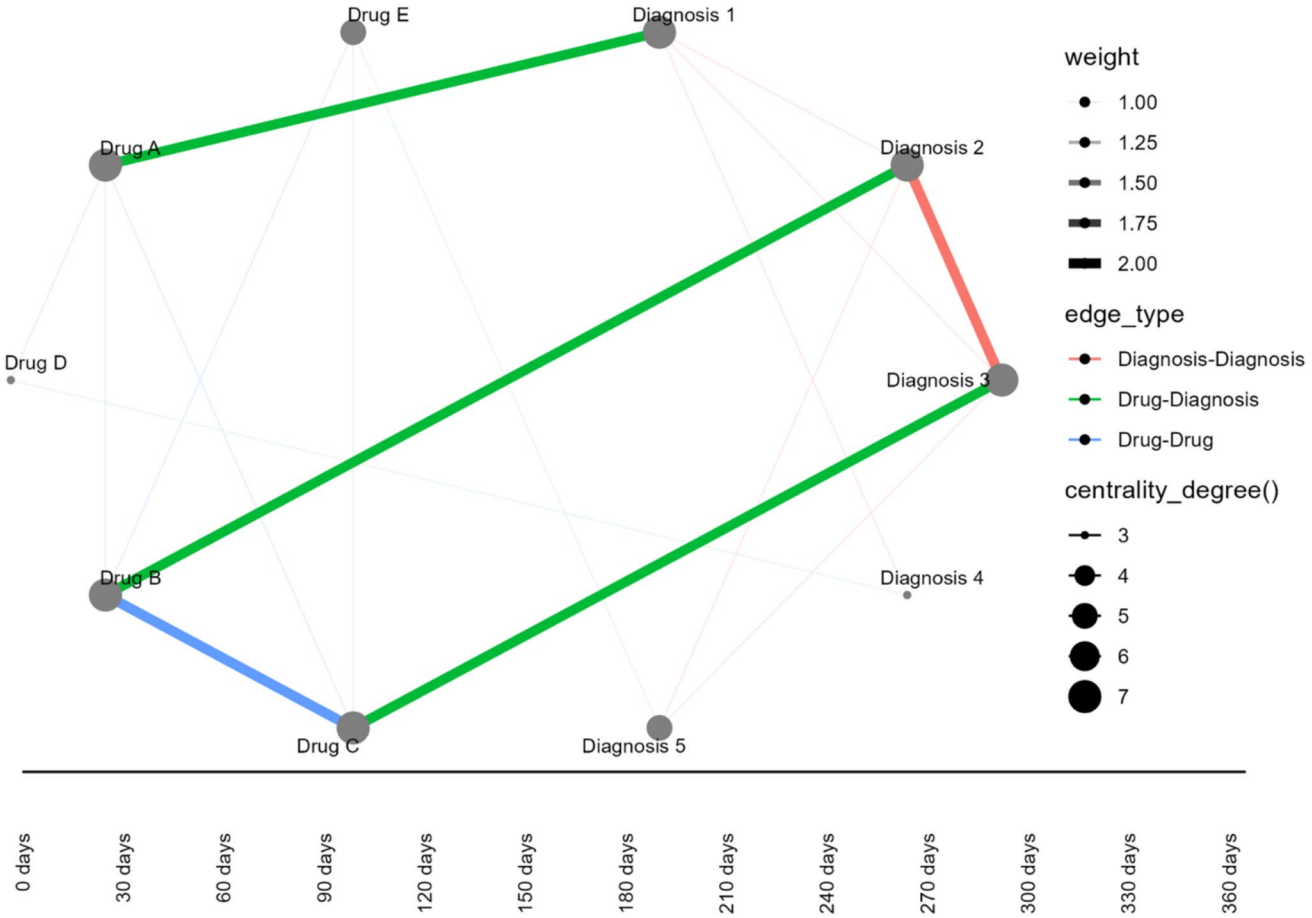

Fig. 2 Example of a network of diagnoses and drugs from 1 year leading up to the first episode of acute myocardial infarction for an individual patient. Each node represents either a drug or a diagnosis. The size of the node indicates its centrality degree, which measures the importance of the node within the network. The edges represent the type of pairs: drug–drug pair (blue lines), drug–diagnosis pair (green lines), and diagnosis–diagnosis pair (red lines). The thickness of the lines indicates the strength of the associations, with thicker lines representing stronger connections. The weight refers to the full edge weight. The centrality degree metric is based on the full network and indicates how central a node is within the entire network. Acute myocardial infarction is not represented as a separate node in this figure; instead, it focuses on the network of diagnoses and drugs leading up to the first episode of acute myocardial infarction

We then searched PubMed for observational or clinical trial studies investigating the association between the drug under investigation and AMI. The search strategy employed the terms 'Drug Name' AND 'myocardial infarction'. If there was at least one study describing the association, we considered the potential safety signal as not novel.

For drugs lacking evidence in PubMed and not listing AMI in the SmPC, we assessed whether they have therapeutic indications for diseases with a very high risk of AMI development (confounding by indication and protopathic bias) [47].

2.6 Signal Prioritization

Prioritization of signals was performed by using the F score, C of AMI (i.e., frequency of the drug–AMI pair), and W_F. These variables underwent standardization by subtracting their mean and scaling them by dividing by their standard deviation. These standardized data were then subjected to k-means clustering, an unsupervised machine learning technique aimed at identifying inherent clusters within the dataset based on similarity in standardized values. The number of clusters (k) was set empirically to 5. A fixed random seed (i.e., 1234) was used to ensure the reproducibility of clustering results across analyses.

For prioritization, the clusters were sorted by descending F score, followed by descending frequency C, and finally by ascending W_F.

2.7 Mathematical Framework for the Network Analysis

In this section, we provide the mathematical framework of the network in our signal detection method. We can describe the patient records as:

$$P_k = \left\{S_{Drk}, S_{Di_k} | \forall_k \in \{1,2,\ldots,N\}\right\},$$

$$S_{\mathrm{Dr}k} = \left\{dr_i | dr_i \in S_{\mathrm{Dr}}\right\},$$

$$S_{\mathrm{Di}_k} = \left\{di_i | di_i \in S_{\mathrm{Di}}\right\},$$

where P_k represents the record of the kth patient out of a total number of N patients, whereas $S_{\mathrm{Dr}k}$ and S_{Di_k} represent the set of unique drugs and diagnoses associated with the kth patient, respectively, that are subsets of all the unique recorded drugs $S_{\mathrm{Dr}k}$ and diagnoses S_{Di_k}. Let us define the co-occurrence (frequency) of a drug–diagnosis pair as:

$$C_{ij} = \sum_{k=1}^{N} \begin{cases} 1, & \text{if} dr_i \in S_{\mathrm{Dr}k}, di_i \in S_{\mathrm{Di}_k} \\ 0, & \text{otherwise} \end{cases}.$$

This can be trivially extended for drug–drug and diagnosis–diagnosis pairs as:

$$C_{ij} = \sum_{k=1}^{N} \begin{cases} 1, & \text{if} dr_i, dr_j \in S_{\mathrm{Dr}k} \\ 0, & \text{otherwise} \end{cases},$$

$$C_{ij} = \sum_{k=1}^{N} \begin{cases} 1, & \text{if} di_i, di_j \in S_{\mathrm{Di}_k} \\ 0, & \text{otherwise} \end{cases}.$$

For a drug $dr_i \in S_{\mathrm{Dr}}$ and a diagnosis $di_i \in S_{\mathrm{Di}}$, the drug co-reporting distribution is given by:

$$\varepsilon_{\mathrm{Dr}_{ij}} = \left\{W_{\mathrm{Dr}_k} | W_{\mathrm{Dr}_k} = n_{\mathrm{Dr}_k} - 1, \text{ in} P_k, \forall_k \in \{1,2,\ldots,N\}, \right.$$
$$\left. \text{if} dr_i \in S_{\mathrm{Dr}_k}, di_j \in S_{\mathrm{Di}_k}\right\},$$

where n_{Dr_k} is the number of drugs reported with patient k. Then, the drug-weighting score is given by taking the 75th percentile of the distribution and adding 0.5.

$$\delta_{\mathrm{Dr}_{ij}} = sort\left(\varepsilon_{\mathrm{Dr}_{ij}}\right),$$

$$W_{\mathrm{Dr}_{ij}} = 0.5 + \delta_{\mathrm{Dr}_{ij}}\left[0.75 \times l_{\mathrm{Dr}_{ij}}\right],$$

where $l_{Dr_{ij}}$ is the length of the distribution vector $\delta_{\mathrm{Dr}_{ij}}$.

Similarly, the diagnosis co-reporting score is:

$$\varepsilon_{\mathrm{Di}_{ij}} = \left\{W_{\mathrm{Di}_k} | W_i = n_{\mathrm{Di}_k} - 1, \text{ in} P_k, \forall_k \in \{1,2,\ldots,N\}, \right.$$
$$\left. \text{if} dr_i \in S_{\mathrm{Dr}_k}, di_j \in S_{\mathrm{Di}_k}\right\}$$

$$\delta_{\mathrm{Di}_{ij}} = sort\left(\varepsilon_{\mathrm{Di}_{ij}}\right),$$

$$W_{\mathrm{Di}_{ij}} = 0.5 + \delta_{\mathrm{Di}_{ij}}\left[0.75 \times l_{\mathrm{Di}_{ij}}\right],$$

where n_{Di_k} is the number of drugs reported with patient k, and $l_{\mathrm{Di}_{ij}}$ is the length of the distribution vector $\delta_{\mathrm{Di}_{ij}}$.

The final predictive score for a drug–diagnosis pair is given by summing their drug-weighting and diagnosis-weighting scores, then dividing the co-occurrence value with the result:

$$W_{\mathrm{F}_{ij}} = W_{\mathrm{Dr}_{ij}} + W_{\mathrm{Di}_{ij}},$$

$$F_{ij} = \frac{C_{ij}}{W_{\mathrm{F}_{ij}}}.$$

For drug–drug and diagnosis–diagnosis scores, we followed the same formulation, with the only difference being that we calculated the elements of the co-reporting distributions by $n-2$ for matching types (e.g., for $\varepsilon_{\mathrm{Di}_{ij}}$ with diagnosis–diagnosis pairs) and by simply n for non-matching types (e.g., for $\varepsilon_{\mathrm{Dr}_{ij}}$ with drug–drug pairs). This corresponds to the theoretical background of counting how many other drugs or diagnoses were reported with a given entity pair.

2.8 Descriptive Analyses

Descriptive analyses were conducted to understand the characteristics of the population, drug use, and hospitalization patterns. These analyses examined the total resident population, the number of individuals who redeemed prescriptions, the total number of prescriptions dispensed, and the types of drugs used (identified by ATC code fifth level). Additionally, information on hospitalizations was analysed, including the total number of hospital admissions, the number of individuals hospitalized, and the variety of ICD-9-CM codes documented. The median number of hospitalizations per patient was also calculated. The analysis of AMI cases identified the number of individuals with at least one AMI hospitalization and the total number of recorded AMI diagnoses. Dispensing data were examined for individuals admitted for AMI, focusing on the number of individuals with at least one prescription filled within 365 days before the first AMI diagnosis and the total number of prescriptions dispensed during that period. Descriptive statistics were employed to characterize the demographics of individuals admitted for AMI, including median age and sex distribution. Additionally, the median number of prescriptions redeemed per year and the median time between redeemed prescriptions and the subsequent AMI diagnosis were calculated.

Further descriptive analyses were conducted to characterize the network, such as the total number of unique nodes, representing both individual drugs and diagnoses, the types of connections (edges) categorized as diagnosis–diagnosis,

drug–diagnosis, and drug–drug. The analysis further investigated the distribution of connections for each node type, examining the median *C* of connections for drugs and diagnoses. Additionally, the strength of connections within the network was assessed using different weighting schemes (i.e., W_F, *F*).

All data are presented as median with quartile 1 and quartile 3 (*Q*1–*Q*3) for continuous variables and as absolute and percentage frequencies for categorical variables. The entire analysis was performed using R Studio software, version 2023.12.1+402 [48].

3 Results

3.1 Descriptive Analysis

During the study period from 2014 to 2018, an average 412,002 individuals were resident in Mantova, Italy. Within this timeframe, 250,943 individuals had at least one redeemed prescription, resulting in a total of 1,5705,028 prescriptions being redeemed, with 774 different drugs (ATC code fifth level) redeemed in the database. A total of 206,414 hospitalizations were registered during the study period: 97,620 individuals had at least one hospitalization and 6057 different ICD-9-CM codes were recorded in the database. The median number of hospitalizations per patient was 1 (*Q*1–*Q*3: 1–3).

A total of 3918 individuals had at least one hospital admission for AMI, and a total of 4686 AMI diagnoses were registered. Within this timeframe, 2866 individuals admitted for AMI had at least one redeemed prescription in the 365 days up to the AMI diagnosis, resulting in a total of 498,591 prescriptions being redeemed. Individuals had a median age of 74 (*Q*1–*Q*3: 64–82) years and were mainly males ($n = 1689$; 58.9%). A total of 582 different drugs were redeemed in the database with a median of 40 (*Q*1–*Q*3: 25–59) redeemed prescriptions for each individual per year. The median time from redeemed prescriptions of all drugs in the 365 days before AMI was 162 (*Q*1–*Q*3: 257–75) days.

3.2 Network Characteristics

The network comprises a total of 2968 unique nodes, considering both individual drugs and diagnoses. The network reveals three main edge types: diagnosis–diagnosis (most frequent, with 529,935 connections), drug–diagnosis (235,380 connections), and drug–drug (102,831 connections). The median *C* of connections (i.e., node degree) for drug nodes was 404 (*Q*1–*Q*3: 194–671), and diagnoses had a median of 380 (*Q*1–*Q*3: 216–664) connections. The median edge weight was 1 (*Q*1–*Q*3: 1–3) (i.e., full degree distribution). The analysis reveals variations in the median weights across different edge categories within the network. The median strength for W_{Dr} connections was 7.5 (*Q*1–*Q*3: 5.5–10.3). W_F had a median of 11.8 (*Q*1–*Q*3: 9–14). *F* score had a median of 0.1 (*Q*1–*Q*3: 0.1–0.3).

3.3 Signal Detection

A total of 331 drugs listing 'myocardial infarction' were initially identified in SIDER. After excluding drugs used exclusively in hospitals, those unavailable commercially or not reimbursed in Italy, the number of drugs from SIDER was reduced to 134. A total of 249 signals were identified (Fig. 3). Among these potential safety signals, 85 (63.4%) were found within the 134 SIDER drugs. Of the remaining 164 drugs identified as potential safety signals and not recorded in SIDER, 24 already listed 'myocardial infarction' in the SmPCs and 60 had protopathic or indication bias, leaving 80 signals for prioritization (Table 1).

3.4 Signal Prioritization

Using a clustering approach, we grouped potential safety signals into five distinct clusters based on three key characteristics: *F* score, W_F, and *C* of occurrence. In all, five potential safety signals received the highest priority as defined in the materials and methods section, 17 classified as priority 2, eight assigned priority 3, 23 categorized as priority 4, and the remaining classified with the lowest priority (Fig. 4).

4 Discussion

To the best of our knowledge, the number of studies using network analysis in EHR data is growing [49, 50], and recent research incorporating administrative databases is now emerging [51, 52]. However, none of these studies have focused on safety signals. Our study presents a novel application of network analysis as a method for signal detection and prioritization in EHRs and administrative databases, specifically investigating drug-induced AMI using data from Mantova, Italy, as a proof of concept. Our approach takes inspiration from the recent work of Pétervári et al. [25], who applied network analysis in the EudraVigilance database.

In our analysis, we recognize that the F score and clustering prioritization in our method can be influenced by the frequency of drug–diagnosis pairs within the dataset. Specifically, more common drug–diagnosis pairs tend to appear more frequently, leading to higher co-occurrence values and, consequently, higher prioritization during network analysis. This effect raises a potential concern, as it could result in a bias toward more common associations,

possibly overshadowing rarer drug–adverse drug reaction (AE) pairs. To address this, we implemented a weighting mechanism, including the 75th percentile adjustment, which helps mitigate the influence of highly frequent pairs. Despite this adjustment, the identification of rare AEs remains inherently challenging because of their limited occurrence in available data. Nonetheless, from a public health perspective, our method is designed to prioritize AEs that are not only more frequent but also more severe, which aligns with the goal of identifying critical health risks that require attention. Therefore, although rare AEs are difficult to detect, our approach effectively targets events that have a greater public health impact, ensuring that the method serves its intended purpose of prioritizing clinically significant drug–diagnosis associations.

A key challenge we encountered was how to define the observation period for each patient. Unlike administrative databases, which track patients prospectively, SRS databases like the FDA Adverse Event Reporting System (FAERS) capture only 'snapshots' of individuals at the time an AE occurs; without the number of total users, it is not possible to compute the incidence of an AE. Our objective was to develop a comparable framework for individuals using administrative databases, capitalizing on the extensive data these sources offer. This methodology was adjusted to incorporate the 75th percentile of patients with a specific diagnosis and their corresponding co-medications. Notably, EHRs and administrative databases include all prescribed medications, whereas individual case safety reports are often incomplete, presenting a significant challenge for SRS database analysis [2, 53]. Indeed, our results illustrate the potential of network analysis to identify complex interactions between medications, clinical events, and patient characteristics, offering an innovative alternative to traditional SRS databases. However, individual case safety reports can include drugs not reimbursed or dispensed at community pharmacies, which are typically missing from Italian EHRs and administrative databases.

Network analysis is a powerful tool for exploring complex relationships within healthcare databases, providing valuable insights that traditional analytical methods may overlook [17, 20, 27]. By modelling patients, treatments (i.e., drugs), and conditions (i.e., diagnoses) as interconnected nodes, network analysis enables the identification of patterns and associations (i.e., edges) across large and diverse datasets. This approach is particularly useful for uncovering hidden connections between comorbidities, tracking disease progression, and understanding the interactions between medications and AEs. Our findings reveal that, during the study period from 2014 to 2018, a substantial number of individuals ($n = 250{,}943$) redeemed prescriptions, leading to over 1.5 million prescriptions, with a diverse range of medications involved (774 unique ATC codes). This extensive data set allows for a comprehensive examination of the relationship between drug use and hospitalization for AMI, highlighting the robustness of the EHRs and administrative databases in capturing real-world patient data. The network analysis produced a substantial framework comprising 2968 unique nodes, which indicates a rich tapestry of connections between diagnoses and medications. The predominance of

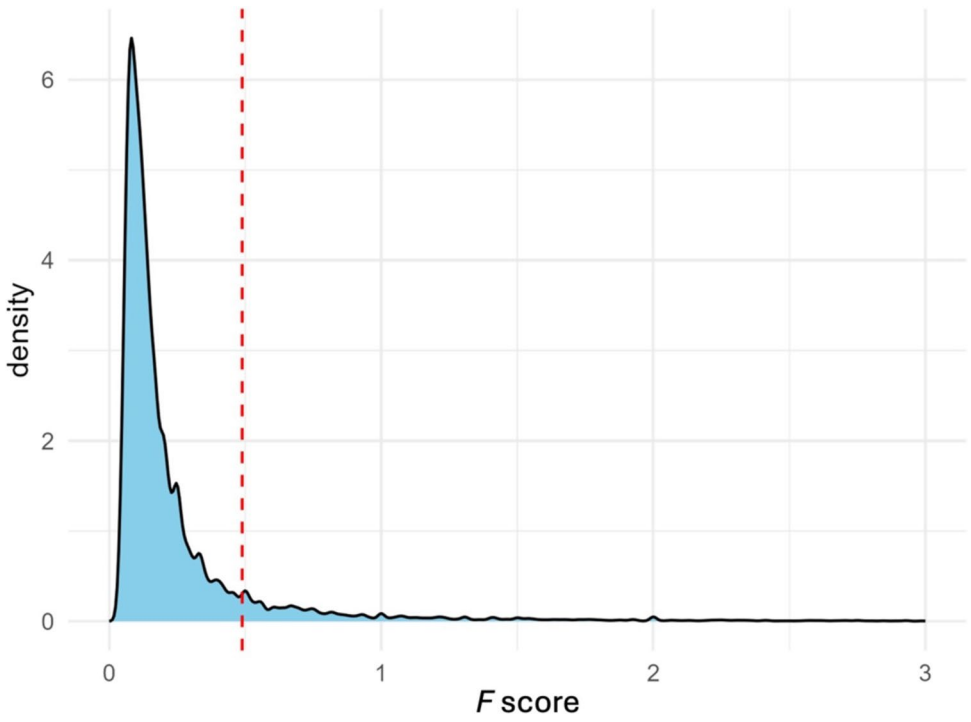

Fig. 3 Signal identification from outliers

Table 1 Potential signals identified from outliers and their prioritization

Drug name	ATC code	Therapeutic indication	Priority (P)
Terazosin	G04CA03	Benign prostatic hyperplasia	P1
Tamsulosin	G04CA02	Benign prostatic hyperplasia	P1
Esomeprazole	A02BC05	Gastrointestinal agent	P1
Omeprazole	A02BC01	Gastrointestinal agent	P1
Allopurinol	M04AA01	Gout treatment	P1
Azithromycin	J01FA10	Antibiotic	P2
Insulin aspart	A10AB05	Antidiabetic	P2
Insulin lispro	A10AB04	Antidiabetic	P2
Insulin lispro (long acting)	A10AC04	Antidiabetic	P2
Alfuzosin	G04CA01	Antihypertensive	P2
Clonidine	C02AC01	Antihypertensive	P2
Dutasteride	G04CB02	Benign prostatic hyperplasia treatment	P2
Silodosin	G04CA04	Benign prostatic hyperplasia treatment	P2
Potassium canrenoate	C03DA01	Diuretic	P2
Torasemide	C03CA04	Diuretic	P2
Alginic acid	A02BX13	Gastrointestinal agent	P2
Ranitidine	A02BA02	Gastrointestinal agent	P2
Ferrous sulfate	B03AA07	Iron supplement	P2
Beclometasone	R03BA01	Respiratory agent	P2
Salmeterol and fluticasone	R03AK06	Respiratory agent	P2
Folic acid	B03BB01	Vitamin B supplement	P2
Calcitriol	A11CC04	Vitamin D supplement	P2
Lincomycin	J01FF02	Antibiotic	P3
Valproic acid	N03AG01	Anticonvulsant	P3
Ebastine	R06AX22	Antihistamine	P3
Haloperidol	N05AD01	Antipsychotic	P3
Calcium carbonate	A12AA04	Calcium supplement	P3
Chlortalidone	C03BA04	Diuretic	P3
Brimonidine	S01EA05	Ophthalmic agent	P3
Selegiline	N04BD01	Parkinson's disease treatment	P3
Oxycodone and naloxone	N02AA55	Analgesic	P4
Cefixime	J01DD08	Antibiotic	P4
Fosfomycin	J01XX01	Antibiotic	P4
Rifaximin	A07AA11	Antibiotic	P4
Sulfamethoxazole and trimethoprim	J01EE01	Antibiotic	P4
Clonazepam	N03AE01	Anticonvulsant	P4
Gliquidone	A10BB08	Antidiabetic	P4
Levocetirizine	R06AE09	Antihistamine	P4
Finasteride	G04CB01	Benign prostatic hyperplasia treatment	P4
Alendronic acid and colecalciferol	M05BB03	Bone metabolism agent	P4
Alendronic acid	M05BA04	Bone metabolism agent	P4
Ibandronic acid	M05BA06	Bone metabolism agent	P4
Methotrexate	L01BA01	Cancer treatment	P4
Multienzymes (lipase, protease, etc.)	A09AA02	Digestive aid	P4
Indapamide	C03BA11	Diuretic	P4
Bimatoprost	S01EE03	Ophthalmic agent	P4
Brinzolamide	S01EC04	Ophthalmic agent	P4
Travoprost	S01EE04	Ophthalmic agent	P4
Montelukast	R03DC03	Respiratory agent	P4
Calcipotriol, combinations	D05AX52	Skin treatment	P4
Clobetasol	D07AD01	Skin treatment	P4

Table 1 (continued)

Drug name	ATC code	Therapeutic indication	Priority (P)
Tacalcitol	D05AX04	Skin treatment	P4
Thiamazole	H03BB02	Thyroid treatment	P4
Codeine and paracetamol	N02AA59	Analgesic	P5
Fentanyl	N02AB03	Analgesic	P5
Ranolazine	C01EB18	Antianginal	P5
Ceftriaxone	J01DD04	Antibiotic	P5
Levetiracetam	N03AX14	Anticonvulsant	P5
Acarbose	A10BF01	Antidiabetic	P5
Insulin aspart	A10AD05	Antidiabetic	P5
Insulin degludec	A10AE06	Antidiabetic	P5
Insulin glulisine	A10AB06	Antidiabetic	P5
Linagliptin	A10BH05	Antidiabetic	P5
Sitagliptin	A10BH01	Antidiabetic	P5
Cetirizine	R06AE07	Antihistamine	P5
Desloratadine	R06AX27	Antihistamine	P5
Rupatadine	R06AX28	Antihistamine	P5
Hydroxychloroquine	P01BA02	Antimalarial	P5
Quetiapine	N05AH04	Antipsychotic	P5
Risedronic acid	M05BA07	Bone metabolism agent	P5
Calcium carbonate + Vitamin D	A12AX	Calcium supplement	P5
Hydroxycarbamide	L01XX05	Cancer treatment	P5
Potassium chloride	A12BA01	Electrolyte supplement	P5
Latanoprost	S01EE01	Ophthalmic agent	P5
Tafluprost	S01EE05	Ophthalmic agent	P5
Budesonide	R03BA02	Respiratory agent	P5
Flunisolide	R03BA03	Respiratory agent	P5
Ipratropium bromide	R03BB01	Respiratory agent	P5
Cyanocobalamin	B03BA01	Vitamin B supplement	P5
Paricalcitol	H05BX02	Vitamin D supplement	P5

diagnosis–diagnosis connections suggests that certain clinical conditions are frequently associated with one another, which is consistent with existing literature on comorbidities [54, 55]. The robust connections between drugs and diagnoses emphasize the need for continued vigilance in monitoring AEs, particularly as patients often receive multiple medications simultaneously [56].

The identification of 249 potential safety signals for AMI, of which 63.4% were linked to drugs already recorded in the SIDER database, underscores the performance of this methodology in surfacing relevant safety information. Notably, the 80 previously unknown signals for drug-induced AMI prioritized for further investigation, particularly those not documented in the SIDER, highlight the importance of ongoing evaluation and re-assessment of drug safety profiles in the context of clinical practice. The clustering approach we used for signal prioritization effectively distilled the safety signals into five distinct clusters based on their characteristics, providing a clear framework for future research. This systematic categorization aids in focusing efforts on the most pertinent signals, enabling healthcare providers and regulatory bodies to allocate resources efficiently toward further investigation and monitoring of potential risks associated with drug use. Our approach successfully prioritized with the highest priority (intended as drugs with the highest mean F score, highest C, and lowest W_F) five drugs: terazosin, tamsulosin, allopurinol, esomeprazole, and omeprazole. These findings are confirmed by previous studies that further corroborate the performance of our method. Proton pump inhibitors (PPIs), such as esomeprazole and omeprazole, are commonly used by the Italian population. A recent study conducted in the Italian Lombardy region, which includes the city of Mantova, identified an elevated risk of cardiovascular events, including AMI, in the elderly population treated with PPIs [57]. Another study [58] hypothesized a potential protopathic bias, suggesting that cardiovascular conditions may present with gastrointestinal-like symptoms; however, chronic use of PPIs is more commonly associated with a diagnosed gastric condition, likely confirmed through imaging tests, or to the primary prevention of peptic ulcers

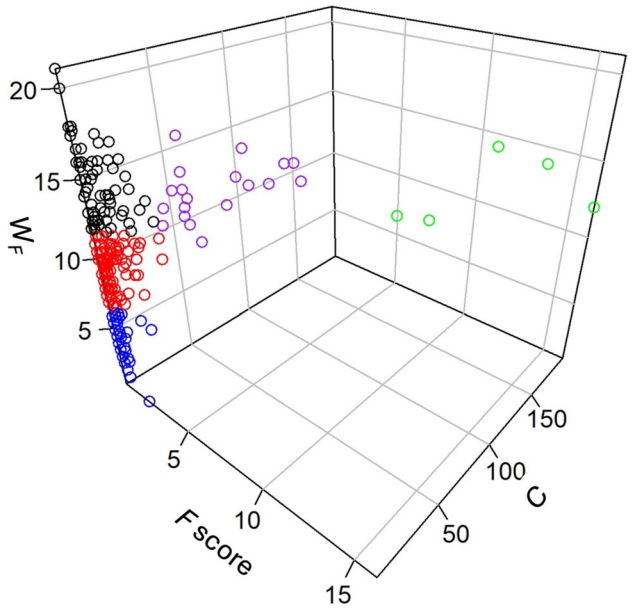

Fig. 4 Signals prioritization. Green = priority 1; purple = priority 2; blue = priority 3; red = priority 4; black = priority 5

in patients taking long-term aspirin for secondary cardiac prevention, rather than to early cardiac manifestations. Given that the impact of concomitant therapies is already accounted for in our analysis, incorporating the duration of PPI exposure could further strengthen our approach. Terazosin and tamsulosin, both alpha-1 adrenergic antagonists primarily used for treating benign prostatic hyperplasia and hypertension, have been associated with adverse cardiovascular effects. These drugs lower blood pressure by causing vasodilation; however, they can also lead to postural hypotension and reflex tachycardia. The cardiovascular stress induced by these side effects may increase the risk of AMI, particularly in older patients or those with pre-existing cardiovascular conditions [59]. Allopurinol, primarily prescribed for gout, has also raised concerns regarding its cardiovascular safety profile. The effectiveness of allopurinol in lowering cardiovascular risk is not yet supported by sufficient clinical evidence, despite its established role in managing uric acid levels [60].

Despite the promising results, this study has limitations that warrant discussion. The reliance on administrative databases may introduce bias due to incomplete data or variations in coding practices. First, the results observed in the population of the city of Mantova may not be applicable at national or international levels given potential regional variations in demographics, healthcare practices, and drug use patterns. Additionally, the sample size was relatively small, which may limit the statistical power and robustness of the conclusions drawn. Moreover, although network analysis offers insights into complex relationships, it may not fully account for confounding variables that influence drug safety, such as patient demographics, comorbid conditions, and clinical practices. Furthermore, although we employed the Pétervári et al. [25] method for signal detection (i.e. calculating various edge weighting scores to be used for network-based signal detection), alternative methodologies exist that could yield different insights. For instance, Fusaroli et al. [26, 61] have published two studies using network analysis for signal detection focusing on the concept of syndromes rather than individual AEs. The study of these clusters is of particular importance in SRS databases, where most of the reports come from consumers. However, we focused our attention on the clinical diagnosis of AMI, which was surely made by a clinician in the hospital. In this case, the symptoms and the general aspect of a more complex syndrome were considered a priori by the reporter. Therefore, our approach could further expand to consider more than one diagnosis to analyse more complex clinical profiles.

Our methods demonstrate potential applicability to other administrative and electronic healthcare databases with comparable structures and data quality. Furthermore, these methods can be extended to investigate other AEs with high positive predictive value and sensitivity, akin to the characteristics of AMI in our study. However, it is important to note that our analysis was limited to a single local health unit, which may constrain the generalizability of our findings. Additionally, the decision to focus on AMI was guided by the original design of the method, which was tailored for application in SRS databases. Future studies should aim to validate these findings across diverse populations and healthcare settings to enhance the generalizability of the results.

5 Conclusions

Our novel method demonstrates that network analysis is a valuable tool for signal detection and prioritization in drug-induced AEs based on EHRs and administrative databases. By leveraging comprehensive RWD, this approach has the potential to enhance pharmacovigilance efforts and improve patient safety outcomes. Continued exploration of network-based methodologies in pharmacovigilance will be crucial for developing a deeper understanding of drug safety in increasingly complex healthcare environments.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40264-025-01515-y.

Declarations

Funding Open access funding provided by Copenhagen University.

Conflicts of Interest MP is the founder and CEO of Sanovigado Kft, a pharmaceutical consultancy and R&D company. PF is the founder and CEO of, and BÁ is employed by, Pharmahungary Group, a group of R&D companies. The work of MaCa is partially supported by the Italian Ministry of Health—Ricerca Corrente—IRCCS MultiMedica. MAB, AA, OMB, VB, MC, AR, CC, ES, and MS have no conflicts of interest to declare.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Material Further inquiries can be directed to the corresponding author.

Code Availability All code access requests can be forwarded to the corresponding author.

Author Contributions MAB: conceptualisation, formal analysis, software, visualisation, writing (original draft), writing (review and editing). AA: formal analysis, visualisation. OMB: mathematical framework. MP: visualisation, writing (review and editing). PF: visualisation. BÁ: visualisation. VB: visualisation, writing (original draft), writing (review and editing). MaCo: visualisation. AR: formal analysis, visualisation. CC: visualisation. MaCa: acquisition of data, supervision, visualisation, writing (review and editing). ES: visualisation. AA: visualisation, writing (review and editing). MS: conceptualisation, methodology, software, supervision, writing (review and editing). All authors read and approved the final version.

References

1. Gauffin O, Brand JS, Vidlin SH, Sartori D, Asikainen S, Català M, et al. Supporting pharmacovigilance signal validation and prioritization with analyses of routinely collected health data: lessons learned from an EHDEN network study. Drug Saf. 2023;46(12):1335–52.
2. Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al. good signal detection practices: evidence from IMI PROTECT. Drug Saf. 2016;39(6):469–90.
3. Caster O, Juhlin K, Watson S, Norén GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. Drug Saf. 2014;37(8):617–28.
4. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. Stat Methods Med Res. 2013;22(1):57–69.
5. Bate A, Hornbuckle K, Juhaeri J, Motsko SP, Reynolds RF. Hypothesis-free signal detection in healthcare databases: finding its value for pharmacovigilance. Ther Adv Drug Saf. 2019;10:204209861986474.
6. Moore TJ, Furberg CD. Electronic health data for postmarket surveillance: a vision not realized. Drug Saf. 2015;38(7):601–10.
7. Coste A, Wong A, Bokern M, Bate A, Douglas IJ. Methods for drug safety signal detection using routinely collected observational electronic health care data: a systematic review. Pharmacoepidemiol Drug Saf. 2023;32(1):28–43.
8. Brown JS, Mendelsohn AB, Nam YH, Maro JC, Cocoros NM, Rodriguez-Watson C, et al. The US Food and Drug Administration Sentinel System: a national resource for a learning health system. J Am Med Inform Assoc. 2022;29(12):2191.
9. Brown JS, Maro JC, Nguyen M, Ball R. Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system. J Am Med Inform Assoc. 2020;27(5):793.
10. Ryan PB, Madigan D, Stang PE, Marc Overhage J, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31(30):4401–15.
11. Ryan PB, Stang PE, Overhage JM, Suchard MA, Hartzema AG, DuMouchel W, et al. A comparison of the empirical performance of methods for a risk identification system. Drug Saf. 2013;36(S1):143–58.
12. Wang SV, Maro JC, Baro E, Izem R, Dashevsky I, Rogers JR, et al. Data mining for adverse drug events with a propensity score-matched tree-based scan statistic. Epidemiology. 2018;29(6):895–903.
13. Trifirò G, Patadia V, Schuemie MJ, Coloma PM, Gini R, Herings R, et al. EU-ADR Healthcare database network vs. spontaneous reporting system database: preliminary comparison of signal detection. Stud Health Technol Inform. 2011;166:25–30.
14. Honig PK. Advancing the science of pharmacovigilance. Clin Pharmacol Ther. 2013;93(6):474–5.
15. Trifirò G, Crisafulli S. A new era of pharmacovigilance: future challenges and opportunities. Front Drug Saf Regul. 2022;2:866898.
16. Aakjær M, De Bruin ML, Andersen M. Epidemiological surveillance of drug safety using cumulative sequential analysis in electronic healthcare data. Basic Clin Pharmacol Toxicol. 2024;134(1):129–40.
17. Askar M, Cañadas RN, Svendsen K. An introduction to network analysis for studies of medication use. Res Soc Adm Pharm. 2021;17(12):2054–61.
18. Askar M. Network analysis: an approach to the study of drug-drug relations. Master's thesis in Pharmacy. University of Norway; 2019.
19. Miglio G, Basso L, Armando LG, Traina S, Benetti E, Diarassouba A, et al. A network approach for the study of drug prescriptions: analysis of administrative records from a local health unit (ASL TO4, Regione Piemonte, Italy). Int J Environ Res Public Health. 2021;18(9):4859.
20. Jeong E, Ko K, Oh S, Han HW. Network-based analysis of diagnosis progression patterns using claims data. Sci Rep. 2017;7(1):15561.
21. Zamora M, Baradad M, Amado E, Cordomi S, Limon E, Ribera J, et al. Characterizing chronic disease and polymedication prescription patterns from electronic health records. In: 2015 IEEE international conference on data science and advanced analytics (DSAA). IEEE; 2015. p. 1–9. Accessed 20 Mar 2024
22. Kim MH, Banerjee S, Zhao Y, Wang F, Zhang Y, Zhu Y, et al. Association networks in a matched case-control design—co-occurrence patterns of preexisting chronic medical conditions

in patients with major depression versus their matched controls. J Biomed Inform. 2018;87:88–95.
23. Hong JC, Hauser ER, Redding TS, Sims KJ, Gellad ZF, O'Leary MC, et al. Characterizing chronological accumulation of comorbidities in healthy veterans: a computational approach. Sci Rep. 2021;11(1):1–11.
24. Monchka BA, Leung CK, Nickel NC, Lix LM. The effect of disease co-occurrence measurement on multimorbidity networks: a population-based study. BMC Med Res Methodol. 2022;22(1):1–16.
25. Pétervári M, Benczik B, Balogh OM, Petrovich B, Ágg B, Ferdinandy P. Network analysis for signal detection in spontaneous adverse event reporting database: application of network weighting normalization to characterize cardiovascular drug safety. Drug Saf. 2022;45(11):1423–38.
26. Fusaroli M, Raschi E, Gatti M, De Ponti F, Poluzzi E. Development of a network-based signal detection tool: the COVID-19 Adversome in the FDA Adverse Event Reporting System. Front Pharmacol. 2021;12: 740707.
27. Brunson JC, Laubenbacher RC. Applications of network analysis to routinely collected health care data: a systematic review. J Am Med Inform Assoc. 2018;25(2):210–21.
28. Cavallo P, Pagano S, Boccia G, De Caro F, De Santis M, Capunzo M. Network analysis of drug prescriptions. Pharmacoepidemiol Drug Saf. 2013;22(2):130–7.
29. Bazzoni G, Marengoni A, Tettamanti M, Franchi C, Pasina L, Djade CD, et al. the drug prescription network: a system-level view of drug co-prescription in community-dwelling elderly people. Rejuvenation Res. 2015;18(2):153–61.
30. Hidalgo CA, Blumm N, Barabási AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol. 2009;5(4): e1000353.
31. Abedian Kalkhoran H, Zwaveling J, van Hunsel F, Kant A. An innovative method to strengthen evidence for potential drug safety signals using electronic health records. J Med Syst. 2024;48(1):1–8.
32. Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. Drug Saf. 2019;42(1):113–22.
33. Garattini L, Badinella Martini M, Zanetti M. The Italian NHS at regional level: same in theory, different in practice. Eur J Health Econ. 2022;23(1):1–5.
34. Franchi C, Ardoino I, Ludergnani M, Cukay G, Merlino L, Nobili A. Medication adherence in community-dwelling older people exposed to chronic polypharmacy. J Epidemiol Community Health (1978). 2021;75(9):854–9.
35. ICD-ICD-9-CM—International classification of diseases, ninth revision, clinical modification. https://www.cdc.gov/nchs/icd/icd9cm.htm#. Accessed 19 Mar 2024.
36. Anatomical Therapeutic Chemical (ATC) Classification. https://www.who.int/tools/atc-ddd-toolkit/atc-classification. Accessed 19 Mar 2024.
37. Franchi C, Ludergnani M, Merlino L, Nobili A, Fortino I, Leoni O, et al. Multiple medication adherence and related outcomes in community-dwelling older people on chronic polypharmacy: a retrospective cohort study on administrative claims data. Int J Environ Res Public Health. 2022;19(9):5692.
38. Casula M, Catapano AL, Piccinelli R, Menditto E, Manzoli L, De Fendi L, et al. Assessment and potential determinants of compliance and persistence to antiosteoporosis therapy in Italy. Am J Manag Care. 2014;20(5):e138–45.
39. Russo V, Orlando V, Monetti VM, Galimberti F, Casula M, Olmastroni E, et al. Geographical variation in medication prescriptions: a multiregional drug-utilization study. Front Pharmacol. 2020;11:1.
40. Mazzali C, Paganoni AM, Ieva F, Masella C, Maistrello M, Agostoni O, et al. Methodological issues on the use of administrative data in healthcare research: the case of heart failure hospitalizations in Lombardy region, 2000 to 2012. BMC Health Serv Res. 2016;16(1):234.
41. Hyeraci G, Spini A, Roberto G, Gini R, Bartolini C, Lucenteforte E, et al. A systematic review of case-identification algorithms based on italian healthcare administrative databases for three relevant diseases of the cardiovascular system: acute myocardial infarction, ischemic heart disease, and stroke. Epidemiol Prev. 2019;43(4 Suppl 2):37–50.
42. Lewer D, Petersen I, Maclure M. The case-crossover design for studying sudden events. BMJ Medicine. 2022;1(1): e000214.
43. Information for myocardial infarction. http://sideeffects.embl.de/se/C0027051/. Accessed 18 Apr 2024.
44. Lists of Class A and Class H medicinal products | Italian Medicines Agency. https://www.aifa.gov.it/en/liste-farmaci-a-h. Accessed 18 Apr 2024.
45. Medicines | European Medicines Agency. https://www.ema.europa.eu/en/medicines. Accessed 20 Mar 2024.
46. Drugs@FDA: FDA-approved drugs. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm. Accessed 20 Mar 2024.
47. Acton EK, Willis AW, Hennessy S. Core concepts in pharmacoepidemiology: Key biases arising in pharmacoepidemiologic studies. Pharmacoepidemiol Drug Saf. 2023;32(1):9.
48. Download RStudio—Posit. https://posit.co/downloads/. Accessed 20 Mar 2024.
49. Martins C, Neves B, Teixeira AS, Froes M, Sarmento P, Machado J, et al. Identifying subgroups in heart failure patients with multimorbidity by clustering and network analysis. BMC Med Inform Decis Mak. 2024;24(1):95.
50. Wan C, Feng W, Ma R, Ma H, Wang J, Huang R, et al. Association between depressive symptoms and diagnosis of diabetes and its complications: a network analysis in electronic health records. Front Psychiatry. 2022;13: 966758.
51. Cavallo P, Pagano S, De Santis M, Capobianco E. General practitioners records are epidemiological predictors of comorbidities: an analytical cross-sectional 10-year retrospective study. J Clin Med. 2018;7(8):184.
52. Longato E, Di Camillo B, Sparacino G, Avogaro A, Fadini GP. Time-resolved trajectory of glucose lowering medications and cardiovascular outcomes in type 2 diabetes: a recurrent neural network analysis. Cardiovasc Diabetol. 2022;21(1):159.
53. Battini V, Cocco M, Barbieri MA, Powell G, Carnovale C, Clementi E, et al. Timing matters: a machine learning method for the prioritization of drug-drug interactions through signal detection in the FDA Adverse Event Reporting System and their relationship with time of co-exposure. Drug Saf. 2024;47(9):895–907.
54. Chudasama YV, Khunti K, Davies MJ. Clustering of comorbidities. Future Healthc J. 2021;8(2): e224.
55. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. Ann Fam Med. 2009;7(4):357.
56. Osanlou R, Walker L, Hughes DA, Burnside G, Pirmohamed M. Adverse drug reactions, multimorbidity and polypharmacy: a prospective analysis of 1 month of medical admissions. BMJ Open. 2022;12(7): e055551.
57. Foresta A, Fernandez LO, Torrigiani G, Schena S, Roncaglioni MC, Nobili A, et al. Proton pump inhibitor use and the risk of cardiovascular complications and death in older adults

with diabetes: a population-based cohort study. Drugs Aging. 2024;41(3):239–49.
58. Chui CSL, Cheung KS, Brown JP, Douglas IJ, Wong ICK, Chan EW, et al. Proton pump inhibitors and myocardial infarction: an application of active comparators in a self-controlled case series. Int J Epidemiol. 2023;52(3):899–907.
59. Garfield S, Judah G, Greta Armando L, Baroetto Parisi R, Remani E, Esiliato M, et al. Persistence to medications for benign prostatic hyperplasia/benign prostatic obstruction-associated lower urinary tract symptoms in the ASL TO4 Regione Piemonte (Italy). Healthcare (Basel). 2022;10(12):2567.
60. Liuzzo G, Patrono C. Allopurinol does not improve cardiovascular outcomes in ischaemic heart disease. Eur Heart J. 2023;44:1016–7.
61. Fusaroli M, Polizzi S, Menestrina L, Giunchi V, Luca P, Raschi E, et al. Unveiling the burden of drug-induced impulsivity: a network analysis of the FDA Adverse Event Reporting System. Drug Saf. 2024;7:1275–92.

Authors and Affiliations

Maria Antonietta Barbieri[1,2] · Andrea Abate[1,2] · Olivér M. Balogh[3,4] · Mátyás Pétervári[3,4,5] · Péter Ferdinandy[3,4,6] · Bence Ágg[3,4,6] · Vera Battini[7] · Marianna Cocco[7] · Andrea Rossi[8,9] · Carla Carnovale[7] · Manuela Casula[8,9] · Edoardo Spina[1] · Maurizio Sessa[2]

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

1 Department of Clinical and Experimental Medicine, University of Messina, 98125 Messina, Italy

2 Department of Drug Design and Pharmacology, University of Copenhagen, Jagtvej 160, 2100 Copenhagen, Capital Region, Denmark

3 Cardiometabolic and HUN-REN-SU System Pharmacology Research Group, Department of Pharmacology and Pharmacotherapy, Semmelweis University, Budapest, Hungary

4 Center for Pharmacology and Drug Research and Development, Semmelweis University, Budapest, Hungary

5 Sanovigado Kft, Budapest, Hungary

6 Pharmahungary Group, Szeged, Hungary

7 Pharmacovigilance and Clinical Research, International Centre for Pesticides and Health Risk Prevention, Department of Biomedical and Clinical Sciences (DIBIC), ASST Fatebenefratelli-Sacco University Hospital, Università Degli Studi di Milano, Milan, Italy

8 Epidemiology and Preventive Pharmacology Service (SEFAP), Department of Pharmacological and Biomolecular Sciences, University of Milan, 20133 Milan, Italy

9 IRCCS MultiMedica, Sesto S. Giovanni, 20099 Milan, Italy