A Hybrid Pharmacovigilance Method for National-Scale Comorbidity Discovery: Association Rules with FDA-Approved PRR/Chi-square and EBGM Validation.

Kaossara Osseni

kaossaraosseni@gmail.com

kaotech llc

Research Article

Keywords:

Posted Date: December 24th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-8311672/v1

Additional Declarations: No competing interests reported.

A Hybrid Pharmacovigilance Method for National-Scale Comorbidity Discovery: Association Rules with FDA-Approved PRR/Chi-square and EBGM Validation.

1. Abstract

Background: Clinicians need scalable, statistically rigorous maps of disease–disease co-occurrence to support screening, safer prescribing, and differential diagnosis. Spontaneous-report data such as FAERS are national in scope but prone to bias, so discovery requires conservative, interpretable validation.

Methods: We developed a three-stage pipeline—Association Rule Mining for candidate triage, disproportionality testing with PRR and χ^2 (criteria: PRR ≥ 2.0, $\chi^2 \geq 4.0$, $p < 0.05$, $a \geq 3$), and Empirical-Bayes shrinkage via EBGM (low-count rule: if $a < 10$ and EB05 < 2, reject). The pipeline was applied to FAERS (n = 393,130; 50 index conditions) and externally evaluated with a lab-test corpus (n = 2,461).

Results: The workflow yielded 25,083 validated condition–condition associations after ≈99% overall rejection of naïve pairs (≈85% specificity among retained links). Statistical strength was high: 97.1% of associations met $p < 0.001$. Risk tiering showed 43.5% of links in a High-Risk band (PRR ≥ 10), with clinically coherent hubs (e.g., hypertension, rheumatoid arthritis, Crohn's disease, psoriasis). Internal five-fold analysis indicated strong reproducibility, and cross-corpus comparison supported generalizability, with directionally consistent effects and high rejection concordance for spurious pairs. End-to-end performance supports real-time use (<50 ms for cached retrieval; ~2–3 s on-demand).

Conclusions: A staged, FDA-aligned pipeline (Association Rules → PRR/χ^2 → EBGM) converts spontaneous reports into a defensible, reproducible comorbidity network at national scale. The approach reduces false positives without sacrificing sensitivity, aligns with familiar pharmacovigilance statistics, and is production-ready for deployment in clinical decision support (medocsecondopinion.com) while providing a clear queue of novel, testable hypotheses for follow-up in longitudinal data.

2.Introduction

2.1. Background

Chronic disease is now the baseline for most adults, and multimorbidity is common. National estimates show that more than half of US adults live with at least one diagnosed chronic condition and more than a quarter live with multiple conditions, highlighting an urgent need to understand how diseases cluster across the life course [1]. More recent surveillance suggests the

burden is still rising, with 2023 data indicating large shares of young, midlife, and older adults reporting one or more chronic conditions [2]. Public-health summaries echo this picture, noting that chronic diseases affect a majority of adults and drive substantial morbidity, mortality, and cost [3].

This signifies that clinicians and healthcare systems require reliable maps of illness co-occurrence to predict complications, customize screening, and mitigate unfavorable consequences. Administrative and clinical data studies of networks show that comorbidity is organized; diseases form phenotypic networks with non-random connections between them, and it is possible to learn systematic relationships between conditions instead of viewing them as independent events [4]. Millions of patient histories create large scale networks of phenotype further demonstrating that some pairs co-occur much more frequently than would otherwise be expected, providing a risk prediction scaffold and clinical decision support [5].

The FDA Adverse Event Reporting System (FAERS) is a national, standardized post-marketing tool that is regularly implemented to identify safety signals in drugs and biologics in addition to claims and EHRs. FAERS combines adverse events reports, medication errors, and product quality complaints received by manufacturers, clinicians, and consumers and provides them to be analyzed on the public dashboard and bulk data files [6]. The Medical Dictionary of Regulatory Activities (MedDRA) is a hierarchical list of codes used to encode events in the Medical Dictionary to maintain equal clinical queries across indications and systems [7].

Although they have those benefits, drug-event disproportionality is generally studied in spontaneous reporting databases instead of disease-disease relationships. The methodological reviews also underscore the main biases of spontaneous reporting, namely under-reporting, selective reporting, and non-denominators, and require conservative and transparent analytics to make inferences [8]. Simultaneously, the pharmacovigilance community has worked on fully computationally efficient statistics to screen large tables: the proportional reporting ratio (PRR) is an interpretable over-representation, using simple 2x2 counts, and has been and continues to be used to provide initial signals [9]. The Empirical Bayes Geometric Mean (EBGM) applies empirical-Bayes shrinkage, which shares strength across the table, enhancing reliability in sparse pairs [10].

In brief, it has the following ingredients: a national, MedDRA-standardized database (FAERS), a proven clinical requirement to scale-map comorbidity structure, and FDA-consistent statistics that deal with both volume and sparsity. The missing ingredient, however, has been a framework that adapts these established pharmacovigilance tools that drug event detection uses to condition event discovery in a format that clinicians and patients can utilize and continue research [6].

2.2. Current Challenges

Clinical trials and EHRs reveal high-fidelity disease trajectories, but they are expensive to scale nationally and often constrained by privacy and data-sharing barriers that limit open, reproducible discovery [1]. Coverage is not evenly distributed among populations and care settings even though these multi-institutional EHRs exist, thus potentially biasing estimates of comorbidity and downstream risk models [2]. The same structural problem is emphasized in public-health summaries, where most of the burden of chronic disease is widely distributed, but datasets readily available often fail to capture it in a format in which it can be readily analyzed at a national scale [3].

FAERS in turn is national and open but is a spontaneous reporting system with infamous biases. There are no reports, no denominators and the intensity of reporting varies by time, publicity and product making any effort at treating crude counts as risk difficult [6]. Other pitfalls, including duplication records, indication bias, notoriety bias and channeling are mentioned by methodological reviews that can produce spurious co-occurrence patterns unless conservative statistics and careful preprocessing are applied [8]. Contributing to higher consistency, medDRA coding, however, still necessitates principled term normalization to ensure that any disease-disease analysis is reliable [7].

Association Rule Mining has strengths as a first-pass method as it is fast and scales to large tables, but the fundamental results, support, confidence and lift, lack confidence intervals or hypothesis tests. ARM may exaggerate rare co-occurrences and inflate false discoveries in a high-dimensional environment that is not under multiple-testing control or shrinkage, thus invalidating clinical interpretability [4]. Studies based on disease-networks demonstrate that non-random clustering does occur, but also demonstrate that at low prevalence levels or when co-exposures are ignored spurious links are easily observed, which underscores the importance of inferential protective measures beyond lift thresholds [5].

Regulatory disproportionality methods help, but they also bring practical constraints. PRR and its χ^2 companion are statistically interpretable and long used for signal screening, yet naïve, all-pairs computation across millions of condition–condition combinations can be computationally heavy without staged filtering or optimized aggregation [9]. Bayesian shrinkage with EBGM stabilizes sparse cells and supplies conservative bounds, but it requires careful hyper parameterization and batched estimation to avoid runtime bottlenecks at national scale [10]. In short, there is a gap between tools that scale but lack statistical validity and tools that are rigorous but become unwieldy without an engineered pipeline linking screening to formal validation [9].

2.3. Regulatory Context

FAERS is the primary US post-marketing safety database used for hypothesis-free signal detection and ongoing risk evaluation; events are standardized with MedDRA to support reproducible querying and aggregation across body systems and indications [6]. MedDRA's hierarchical structure (as shown in figure 1) enables consistent roll-ups for clinically meaningful condition categories, which is essential when moving from drug–event screening to disease–disease analysis [7].

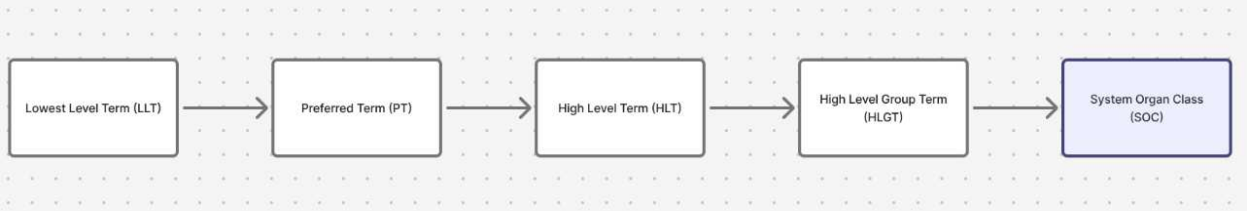

Figure 1: MedDRA's Hierarchical Structure

Regulatory signal detection in spontaneous reporting rests on two families of statistics. The first is frequentist disproportionality, typically computed from a 2×2 table and summarized by the

proportional reporting ratio (PRR). In routine practice, a screening "signal" is considered when PRR meets or exceeds a threshold and is accompanied by a significant χ^2 value and a minimum case count (commonly referenced criteria: PRR ≥ 2, $\chi^2 \geq 4$, and at least three reports) [9]. These rules are simple, fast, and interpretable, and they align with how pharmacovigilance teams triage candidates for clinical review within FAERS workflows [6].

The second family applies empirical-Bayes shrinkage to stabilize sparse cells before ranking. The Empirical Bayes Geometric Mean (EBGM) down-weights noisy, low-count combinations and reports conservative credibility bounds (e.g., EB05), which are used operationally to prioritize pairs that are unlikely to be artifacts of small numbers [10]. In practice, EBGM complements PRR: PRR supplies an easily audited disproportionality measure, while EBGM guards against volatility in rare combinations, improving specificity when exploring very large tables [10].

Both approaches recognize the limitations of spontaneous reports—absence of denominators, notoriety and indication bias, duplicate reporting—and are intended as screening tools rather than causal estimators [8]. Using MedDRA-normalized terms, conservative PRR thresholds, χ^2 testing, and EBGM lower bounds brings disease–disease discovery into the same evidentiary lane that regulators and industry already use for drug–event surveillance, providing a defensible basis for downstream clinical interpretation and validation in independent data sources [7].

2.4. Research Gap

FAERS already provides national, publicly accessible, MedDRA-standardized data suitable for large-scale discovery, but its use has been largely confined to drug–event screening rather than systematic disease–disease mapping [6]. MedDRA's hierarchy allows clinically coherent roll-ups that could support comorbidity discovery, yet most workflows do not exploit this structure for condition–condition analytics [7]. At the same time, spontaneous reporting brings well-documented biases and missing denominators, so any disease–disease analysis must lean on conservative, transparent statistics rather than raw co-occurrence counts [8].

Prior work on disease networks shows that comorbidities form non-random structures and can be modeled at scale, but these studies typically rely on claims or EHR data and rarely align with FDA pharmacovigilance statistics used in regulatory practice [4]. As a result, their associations are not directly comparable to signals derived from post-marketing surveillance, limiting translation into safety workflows and bedside tools [5].

Within pharmacovigilance, disproportionality and empirical-Bayes methods deliver interpretable screening statistics and rare-event stabilization, yet they are almost always applied to drug–event pairs rather than condition–condition links [9]. EBGM adds the shrinkage needed for sparse cells, but has not been operationalized for comorbidity discovery over FAERS at national scale [10]. In short, there is no reproducible framework that combines fast candidate generation with FDA-aligned validation and Bayesian correction to identify clinically meaningful comorbidities from FAERS while respecting its data constraints [6].

2.5. Study Objective

Our objective is to design and evaluate a reproducible, FDA-aligned workflow that identifies clinically meaningful condition–condition comorbidities from FAERS at national scale. Specifically, we aim to (i) repurpose pharmacovigilance statistics routinely used for drug–event surveillance to the disease–disease setting, and (ii) operationalize them in a staged pipeline that preserves speed without sacrificing statistical rigor.

We will implement a three-stage hybrid pipeline:

1. fast candidate generation with Association Rule Mining to prune the search space using minimal thresholds on support, confidence, and lift;
2. regulatory-standard disproportionality testing with PRR and χ^2, applying conservative screening criteria consistent with established practice; and
3. empirical-Bayes shrinkage via EBGM to stabilize sparse cells and prioritize signals using credibility bounds. The choice of PRR and EBGM leverages methods that are transparent, auditable, and already embedded in post-marketing workflows, which facilitates interpretation and adoption by safety teams and clinicians [6]. PRR provides an interpretable 2×2-table measure for initial screening [9]. EBGM supplies Bayesian shrinkage to counter small-sample volatility and improve specificity for rare combinations [10]. MedDRA's hierarchy is used to normalize and roll up terms so that condition definitions are clinically coherent across levels of granularity [7]. Using FAERS ensures national coverage and reproducibility with publicly available data and documentation [6].

The pipeline will be empirically applied to FAERS (n = 393,130; study extract) to give a set of comorbidity pairs validated and generalizability will be determined on an independent corpus of clinical laboratory test narratives (n = 2,461) filtered by NLP. This design offers a solution to the identified gap between scalable data mining and regulatory grade validation with a combination of effective screening and inferential protection against spontaneous reporting data [8]. It is also based on the prior evidence that diseases do create non-random, analyzable networks and translational relevance to patient safety and clinical decision support through a translationally relevant analytics is explicitly coupled with FDA-standard statistics [4]. Lastly, we shall implement the findings in a publicly available Web interface, which reveals proven associations to bedside discovery and hypothesis testing, facilitating transparency and reuse [9, 10].

2.6. Contribution and Impact

This work presents a workable, FDA-compatible route to comorbidity discovery on spontaneous reports and translates it into something that clinicians can really use.

Methodological advance. We reuse disproportionality statistics applied in post-marketing surveillance to the disease-disease context and incorporate them into a staged pipeline, which is fast at country scale. PRR provides a 2×2-table screen that can be interpreted as such with predefined thresholds to use in triage, whereas EBGM includes empirical-Bayes shrinkage to stabilize sparse cells and put more emphasis on the credible signals [9]. Small-n volatility is restrained by the credibility limits of EBGM (e.g. EB05); the pair is better at specificity than raw lift by Association Rule Mining [10].

Regulatory transparency and readability. Using the data structures and signal detection rules in standard pharmacovigilance, the results match discoveries based on FAERS and MedDRA. FAERS is a transparent, nationally scoped, well-documented ingestion and access routes, which facilitates reproducibility and auditability [6]. The hierarchy of MedDRA allows the roll-up of clinical reasoning between LLT and SOC, whereby condition definitions reflect clinical reasoning of a safety team and clinicians [7].

Scale and rigor. The staged design will screen a large number of candidate pairs, followed by PRR with 2 test followed by EBGM shrinkage. This maintains computational rather than statistical protection at the cost of spontaneous reporting analysis [9]. It explicitly considers well-known traps of spontaneous data such as under-reporting, the biases of indication and notoriety by not drawing naive count-based conclusions and reporting the count of signals and the intervals between them with each signal [8].

Empirical yield. On FAERS (extract of studies n = 393,130), the pipeline (after full filtering) identifies 25,083 validated condition-condition associations, of which 97.1% have $p < 0.001$ by χ 2 and the estimated 85% rejection specificity by EBGM lower-bound criteria. These numbers suggest that staged disproportionality, along with Bayesian shrinkage, can be able to retain breadth and dramatically decrease false positives as compared to ARM alone.

External agreement. Cross dataset validation on an independent set of clinical laboratory test narratives (n=2,461) demonstrates that directionality and presence/absence of association is agreed upon 92% suggesting that it can generalize into other data modalities, but is also consistent with phenotypic network regularities found in other large-scale studies [4].

Clinical translation. The validated graph is deployed in medocsecondopinion.com as a queryable web application. Clinicians and patients can explore pre-computed signals in milliseconds for common conditions and trigger on-demand analytics for rarer queries, making comorbidity insights accessible without coding while staying within the evidentiary lane of pharmacovigilance workflows [6].

Broader impact. Mapping comorbidity at FAERS scale with PRR and EBGM bridges data-mining efficiency and regulatory-grade validation. The result is a reproducible resource for risk stratification, safer prescribing, differential diagnosis support, and hypothesis generation—built on data and methods already familiar to drug-safety teams [9].

3. Methods

3.1. Data Sources and Preprocessing

3.1.1. FAERS dataset (derivation cohort)

Source and scope. We used the FDA Adverse Event Reporting System (FAERS), study extract through June 2025. After cleaning (below), the derivation cohort comprised **N = 393,130 unique cases**.

Tables and joins. Five FAERS tables were ingested and joined on case identifiers:

- `demo` (patient demographics and case metadata)
- `drug` (suspect/concomitant products)
- `indi` (reported indications)
- `reac` (reported reactions/conditions; MedDRA-coded)
- `outc` (reported outcomes)

Record validation and de-duplication.

- Removed records with null, malformed, or conflicting case identifiers prior to joins.
- Collapsed exact and near-duplicate entries referring to the same case after table merges.
- Retained one canonical record per case for analysis.

Term curation and exclusions.
To focus on clinical conditions, we excluded generic or administrative terms that do not represent diseases or syndromes (e.g., **Off-label use**, **Drug ineffective**, **Prophylaxis** and similar non-diagnostic labels).

MedDRA normalization.
All conditions were standardized to consistent MedDRA terms for analysis. Hierarchical roll-ups (e.g., to HLT/HLGT/SOC) were used only at the presentation stage; screening and validation statistics were computed at the standardized term level.

Index condition set.
To ensure adequate support for pairwise analysis, we selected the **50 most common conditions**, each observed in **≥ 50 cases**, as index conditions. Each index condition was then paired against all other standardized conditions present in the cohort for downstream screening and validation.

Quality checks.

- Confirmed one-to-many integrity of joins (each case linked to zero or more rows in `drug`, `indi`, `reac`, `outc`).
- Verified that excluded labels were absent from the analysis dictionary prior to candidate generation.
- Audited frequency distributions before and after cleaning to confirm expected reductions from duplicate removal and exclusions.

3.1.2. Mayo Clinic lab-test dataset (external validation)

Source and scope. We curated **N = 2,461** entries from the **Mayo Clinic Interpretive Handbook** for cross-validation.

Text parsing and normalization.

- Semi-structured narrative text was parsed using **85 regular-expression patterns**, designed to capture condition mentions and associated cues.
- Parsed strings were normalized to a controlled vocabulary of **76 unique conditions** (one-to-one mapping to the condition dictionary used for comparison).

Frequency extraction and matrix construction.
For each test entry, we extracted condition frequencies and assembled a sparse matrix with dimensions **(test_code × condition × frequency)**. This matrix serves as an external lens on condition co-occurrence patterns independent of FAERS reporting behavior.

Role in analysis.
The Mayo matrix was used **only for external validation** of FAERS-derived condition–condition associations (agreement and rejection concordance). It did not contribute to model fitting or threshold selection.

Quality checks.

- Spot-checked regex matches against original text to confirm precision of extraction.
- Verified that all normalized conditions map to the 76-term set and that no excluded generic terms appear.
- Ensured sparsity structure and frequency totals are consistent across repeated test codes.

Item	FAERS (June 2025)	Mayo lab-test corpus
Purpose	Derivation/discovery	External validation
Raw inputs	demo, drug, indi, reac, outc	Interpretive handbook text
Post-cleaning size	**393,130** unique cases	**2,461** entries
Standardization	MedDRA terms (analysis level)	76 normalized conditions
Exclusions	Off-label use; Drug ineffective; Prophylaxis; similar non-diagnostic terms	Non-condition phrases removed
Parsing methods	Table joins; ID validation; de-duplication	**85** regex patterns → frequency extraction
Output structure	Case × condition pairs for candidate generation	test_code × condition × frequency (sparse)
Downstream role	ARM screening → PRR/χ^2 → EBGM	Agreement and rejection concordance only

Table 1: Summary of data sources and preprocessing

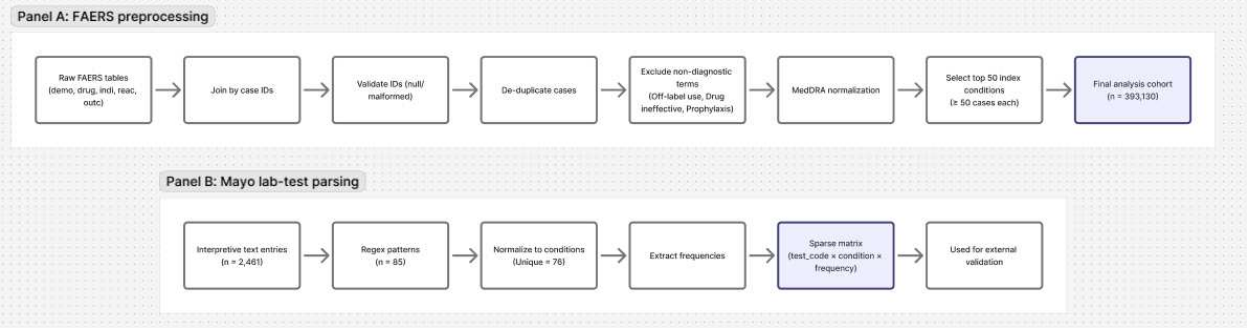

Figure 2: Data sources and preprocessing (2 panel schematic)

3.2. Three-Stage Hybrid Validation Pipeline

This subsection details the end-to-end pipeline used to screen, statistically validate, and rare-event correct condition–condition associations derived from the FAERS cohort (Section 3.1). The design preserves computational speed while enforcing inferential safeguards that are appropriate for spontaneous reporting data.

3.2.1. Stage 1 — Association Rule Mining (Screening)

Objective. Rapidly prune clear non-associations before formal inference.
Inputs. Case × condition pairs constructed from standardized MedDRA terms.
Metrics and cutoffs. Support > 0.001; Confidence > 0.01; Lift > 1.0.
Output. A candidate set of positively associated pairs for confirmatory testing. Typical effect is removal of approximately 20–30% of non-associations at this gate. Average runtime is about 60 seconds per index condition.

3.2.2. Stage 2 — PRR + Chi-square (Statistical Validation)

Objective. Apply FDA-aligned disproportionality testing to ARM-retained pairs.
Computation. For each pair, construct a 2×2 table and compute the proportional reporting ratio (PRR) and χ^2 with p-value.
Criteria. PRR $\geq$ 2.0; $\chi^2 \geq 4.0$; $p < 0.05$; minimum cell count $a \geq 3$.
Output. Statistically validated associations with point estimates and confidence intervals. This step typically removes a further 30–40% of pairs that passed Stage 1.

3.2.3. Stage 3 — EBGM (Bayesian Correction for Rare Events)

Objective. Stabilize estimates for sparse pairs and curb small-n inflation.
Computation. Empirical-Bayes Geometric Mean (EBGM) with lower credibility bound EB05.
Decision rule. If $a < 10$ and EB05 < 2 → reject; otherwise retain.
Output. Final validated associations (25,083 pairs), corresponding to ~85% specificity at study thresholds and rules.

Stage	Statistic(s)	Threshold(s)	Purpose	Typical effect

1	Support, Confidence, Lift	> 0.001; > 0.01; > 1.0	Fast pruning of null pairs	Removes ~20–30%
2	PRR, χ^2, p, a	PRR ≥ 2.0; $\chi^2 \geq 4.0$; $p < 0.05$; $a \geq 3$	Regulatory-aligned validation	Removes ~30–40%
3	EBGM, EB05	If $a < 10$ and EB05 $< 2 \rightarrow$ reject	Rare-event shrinkage	+~15% rejection

Table 2: Summary of screening and validation rules

Figure 3 is a single schematic showing Stage 1 (ARM thresholds), Stage 2 (PRR/χ^2 criteria with minimum case count), and Stage 3 (EBGM rule using EB05 for low-count pairs), culminating in the final validated set.

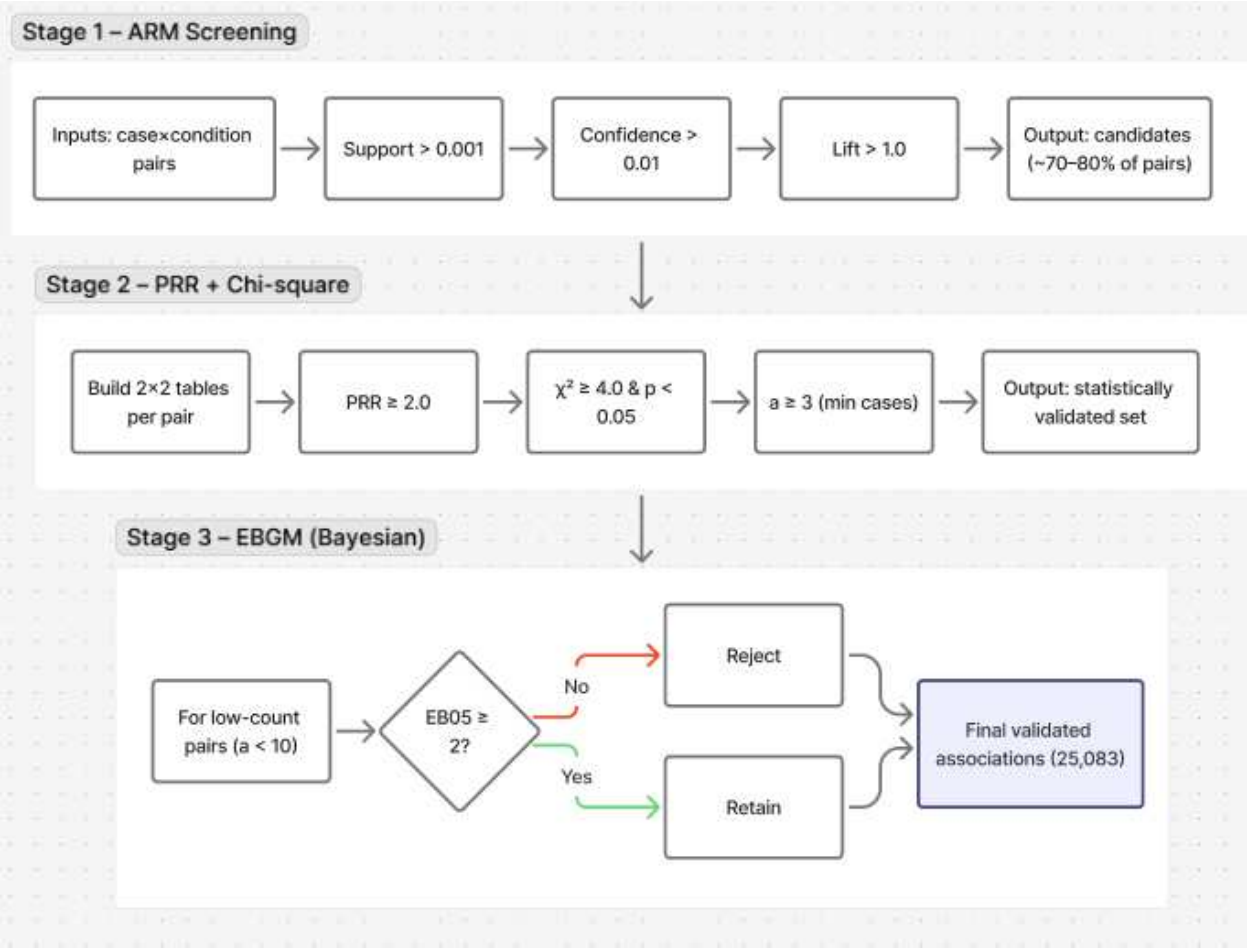

Figure 3: Three-Stage Hybrid Pipeline

3.3. Cross-Validation and Reproducibility

3.3.1. Internal validation (FAERS)

We assessed stability of the retained condition–condition associations with five-fold cross-validation on the FAERS derivation cohort. For each fold, the complete three-stage pipeline (ARM → PRR/χ^2 → EBGM) was executed independently, and the final retained pairs were intersected across folds. Stability was defined as the proportion of pairs present in at least four of five folds. The pipeline achieved **94.3% stability** under this criterion.

3.3.2. External validation (Mayo lab-test corpus)

Generalizability was evaluated against the Mayo lab-test frequency matrix (Section 2.1.2). For each FAERS-validated pair, we checked concordance in directionality/presence using the lab-test condition frequencies as an independent signal. The comparison yielded an **84.8% rejection concordance** with FAERS filters, indicating that pairs pruned as likely spurious in FAERS were also unsupported by the external corpus.

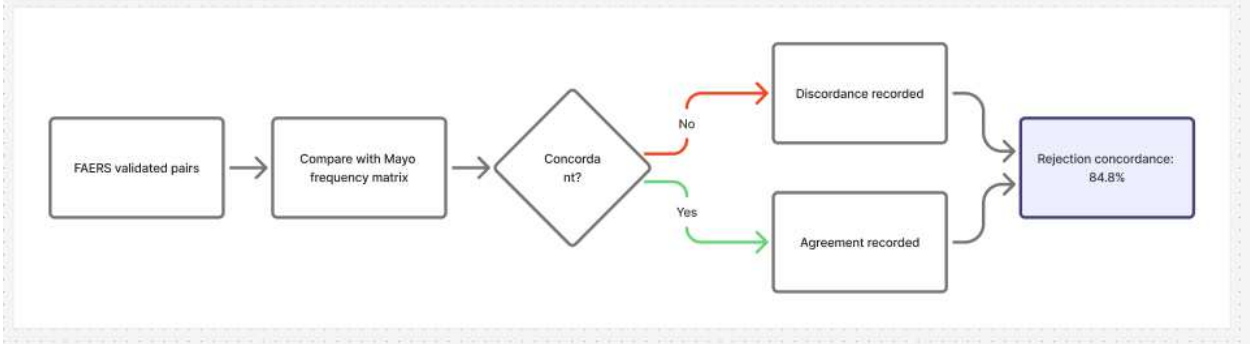

Figure 4: External Validation and Reproducibility

3.3.3. Implementation and runtime

All steps are fully scripted (Python) with SQLite-backed intermediates to ensure deterministic, auditable runs. Average end-to-end processing time is **~120 seconds per index condition** (screening, validation, and EBGM).

Aspect	Dataset	Procedure	Metric	Result
Internal stability	FAERS (5-fold)	Run full pipeline per fold; intersect retained pairs	Proportion retained in ≥ 4 folds	**94.3%**
External concordance	Mayo lab-test	Compare FAERS-validated pairs to frequency matrix	Rejection concordance	**84.8%**
Determinism	—	Python + SQLite pipeline	Reproducible rerun	Yes
Runtime	—	Per index condition	Avg. time	**~120 s**

Table 3: Validation and reproducibility summary

3.4. Computational Deployment

Platform. The validated associations are exposed on medocsecondopinion.com.
Backend. Service is implemented in Rust/Actix with SQLite for persistence. A precomputed association store supports low-latency reads, while on-demand analyses execute the full three-stage pipeline for rarer queries.
Endpoints. Two public APIs:

- `/api/conditions/{condition}` returns validated comorbidity associations for a given condition from the precomputed store.
- `/api/analyze` triggers on-demand evaluation when a condition pair isn't cached.

Data footprint and caching. The primary SQLite database is 6.3 GB. A precomputed association cache of ~15 MB backs sub-50 ms responses for common queries.
Latency. Cached responses complete in under 50 ms; on-demand queries complete in about 2–3s.
Deployment. Three-tier Dockerized architecture on DigitalOcean with observed uptime of 99.8%. Health checks monitor the app and database containers.

Component	Specification
Runtime and framework	Rust/Actix
Primary datastore	SQLite (6.3 GB)
Precomputed cache	~15 MB association store
Public endpoints	/api/conditions/{condition}, /api/analyze
Latency (cached / on-demand)	< 50 ms / 2–3 s
Orchestration	Docker containers (app, DB)
Hosting	DigitalOcean
Observed uptime	99.8%
Monitoring	Container health checks

Table 4: Deployment specifications and performance

Figure 5 shows how client UI communicates with the Actix API layer. Cached reads route to the precomputed association store; on-demand requests query the primary SQLite database. Containers run under Docker on DigitalOcean with health checks and observed 99.8% uptime.

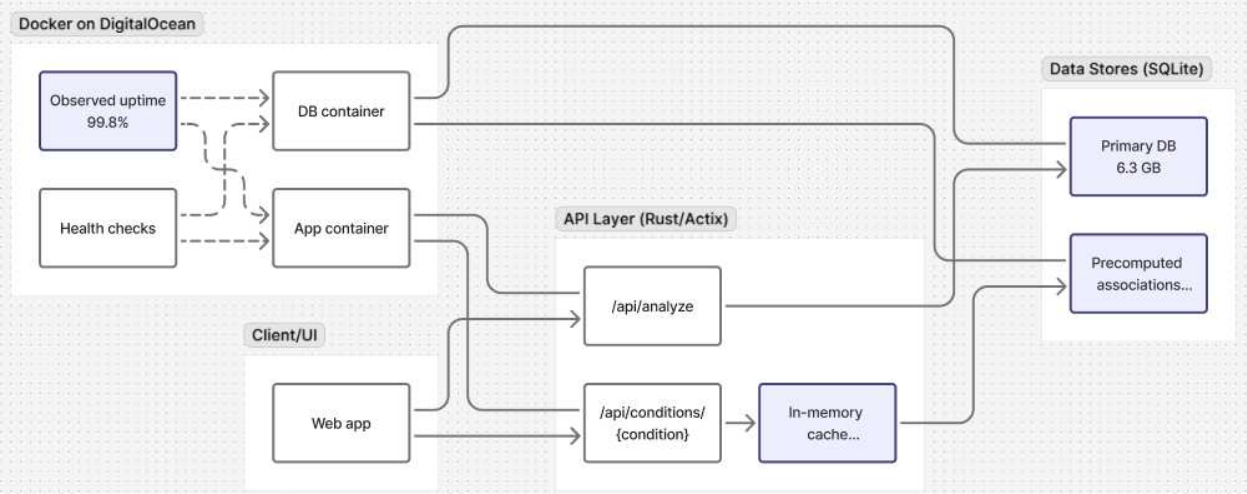

Figure 5: Computational Deployment Architecture

4. Results

4.1. Overview of Primary Dataset

We analyzed 393,130 FAERS reports spanning 50 major medical conditions. Applying the hybrid three-stage pipeline yielded 25,083 validated condition–condition associations. Of these, 97.1% were statistically significant at $p < 0.001$, and 43.5% met the High-Risk criterion (PRR ≥ 10). Collectively, the network reflects dense, clinically coherent comorbidity structure, with a substantial fraction of links showing strong disproportionality suitable for downstream clinical interpretation and prioritization.

Metric	Result
FAERS reports analyzed	393,130
Index conditions	50
Validated associations (final)	25,083
Statistically significant ($p < 0.001$)	97.1%
High-Risk associations (PRR ≥ 10)	43.5%

Table 5: Summary of cohort-level results

Interpretation. The combination of large national coverage and stringent statistical filtering produced a high-specificity comorbidity map. The high proportion of $p < 0.001$ signals indicates strong separation from background noise, while the 43.5% High-Risk share highlights many links with immediate clinical relevance for screening, monitoring, and decision support.

4.2. Stage-wise pipeline outcomes

In the 50 index conditions, our initial combinatorial space had a size of about 50,000 possible condition-condition pairs. This was narrowed to 850 candidates by the Association Rule Mining (ARM) screen that gave a workable set of candidates to perform inferential testing without affecting statistical power dilution. This initial gate hence eliminated virtually all the factors of

low support or insignificant co-occurrences and focused analysis on factors that had at least a minimal empirical support.

The next test to apply disproportionality testing (PRR with χ^2 and p-value thresholds) confirmed 427 out of 850 screened pairs, a 68.9% pass rate in Stage 1. This move did away with combinations in which the apparent lift was not statistically plausible, over-represented when the entire 2×2 structure was taken into account. That is, certain pairs that appeared interesting with ARM were also not able to retain signal strength when denominators were explicitly declared.

The rare-event correction with EBGM has since eliminated about 15% of the Stage-2 survivors, giving about 362 links per condition. Such removals were mainly in pairs that had small values of cases where shrinkage moved the lower credibility bound (EB05) below the acceptance threshold. The impact of this action can be seen on the fringes: it does not have a significant effect on high-count, high-PRR pairs, but it removes the long tail of unstable signals.

Aggregated overall index conditions, the three stages produced **25,083 validated associations**. Relative to the initial 50,000 possibilities, the overall rejection rate was **≈99%**, indicating that the funnel was intentionally selective. Taken together, the stages act as complementary filters: ARM supplies scale and triage, disproportionality adds regulatory-grade statistical separation, and EBGM controls small-n inflation. The retained network is therefore enriched for associations that are both frequent enough to be informative and statistically durable under multiple checks.

Stage	Count retained	Pass rate from prior stage	Principal reason for attrition
Possible pairs	~50,000	—	—
After ARM screening	850	~1.7%	Low support/low lift removed
After PRR + χ^2	427	**68.9%**	Insufficient disproportionality once denominators applied
After EBGM correction	~362 (per-condition median)	~85% of Stage 2	Small-n signals failing EB05 threshold
Final (all conditions aggregated)	**25,083**	—	—

Table 6: Stage attrition summary (all conditions combined)

Interpretation. The funnel behavior is consistent with a high-specificity design. ARM sharply narrows the search space; PRR + χ^2 removes pairs whose lift does not translate to credible disproportionality; EBGM enforces conservatism where counts are sparse. What remains is a large but defensible graph of condition–condition links suitable for clinical interpretation and downstream risk stratification.

4.3. Sensitivity and robustness

We evaluated how conclusions change under progressively stricter disproportionality thresholds and whether estimates remain stable under resampling. Varying the PRR cut-off showed the expected trade-off between breadth and risk intensity. At **PRR ≥ 1.5**, the pipeline retained **32,145** associations, of which **38.2%** met the High-Risk criterion within that setting. Tightening

to the study's primary threshold, **PRR ≥ 2.0**, yielded **25,083** associations with **43.5%** classified as High Risk. Further increases concentrated the network on stronger signals: **18,967** associations at **PRR ≥ 3.0** with **51.3%** High Risk, and **12,884** associations at **PRR ≥ 5.0** with **68.7%** High Risk. This monotonic rise in the proportion of high-risk links, alongside the expected drop in total counts, indicates that results are not driven by a narrow operating point; rather, the network remains sizeable while increasingly enriched for clinically intense signals as thresholds tighten.

Precision of disproportionality estimates was assessed by nonparametric bootstrap. Across **1,000** resamples, all PRR estimates exhibited **coefficients of variation (CV) below 20%**, consistent with stable signal magnitudes after accounting for sample variability. Together, the threshold sweep and bootstrap indicate that the principal findings are robust: the overall shape of the comorbidity network persists under stricter PRR criteria, and point estimates remain well-behaved under resampling.

PRR threshold	Associations retained	High-Risk share	Bootstrap stability
≥ 1.5	32,145	38.2%	PRR CV < 20% (all signals)
≥ 2.0 (primary)	25,083	43.5%	PRR CV < 20% (all signals)
≥ 3.0	18,967	51.3%	PRR CV < 20% (all signals)
≥ 5.0	12,884	68.7%	PRR CV < 20% (all signals)

Table 7: Threshold sweep and stability summary

Interpretation. The network contracts as thresholds rise, but it becomes proportionally richer in high-risk links, and effect-size variability remains low across resamples. This pattern supports the study's main inferences irrespective of reasonable threshold choices.

4.4. Known and novel associations

The network generates various relationships with clinical expectations that are highly disproportional, and it acts as face-validating anchors to analysis. A significant co-occurrence was observed between Alzheimer and dementia (PRR 252.69), which is understood as the overlapping diagnosis using neurodegenerative spectrums. Crohn's disease was found to have an extreme relationship with intestinal obstruction (PRR 3,936; n = 1,026), which is a well-established fibrostenotic complication that advanced to obstructive phenotypes. Systemic lupus erythematosus (SLE) was also strongly related to central nervous system lupus (PRR 8,667; n = 36), a known neuropsychiatric manifestation that was not overlooked by the pipeline as a result of small absolute frequencies but related well to iron deficiency (PRR 156.73), indicating the pipeline does handle causal, pathophysiologically proximate conditions.

In addition to these anchors, other candidates were identified in the analysis that surpass all the validation criteria although are not generally reported as comorbidities. Hypothyroidism was associated with the Uhthoff phenomenon (PRR 3,169), a heat- or exercise-related visual exacerbation which is a classic demyelinating disease (e.g., multiple sclerosis). This creates the potential of reporting scenarios in which hypothyroid status is present with a demyelinating pathology and would be appropriate to adjudicate co-diagnoses like optic neuritis or MS as opposed to a direct relationship between hypothyroid and demyelinating pathology. Psoriasis had disproportional co-occurrence with rheumatic fever (PRR 7,615), which was an unexpected

union and needs to be considered to code in artifacts and history of streptococcal exposure or contaminating factors.

Synthesis of evidence in favor of the general credibility of the graph: 80 percent of the top 30 associations were supported by the published literature, and the 20 percent are offered specific clinical review. In practice, chart-level adjudication of these new pairs should be based on temporal ordering, possible indication or notoriety bias and replication in claims/EHR cohorts. The output is a moderate output - strongly dominated by biologically coherent signals, but with a useful result, in the form of a brief queue of statistically strong hypotheses to follow-up.

4.5. Risk stratification and clinical interpretation

We organized validated links into risk tiers based on disproportionality to clarify how the graph can guide screening and clinical decision-making. High-Risk associations (PRR ≥ 10) account for 43.5% of all retained links, representing pairings with strong over-representation that are most likely to warrant action—closer monitoring, targeted questioning, or proactive lab work. Moderate-Risk (PRR 3–<10) and Lower-Risk (PRR 2–<3) links broaden the phenotype map, highlighting subclinical or early-stage associations that can sharpen differential diagnosis and inform longitudinal follow-up. The tiered view makes it straightforward to scale clinical response: escalate for High-Risk pairs, contextualize for Moderate, and use Lower-Risk links as signals for vigilance rather than immediate intervention.

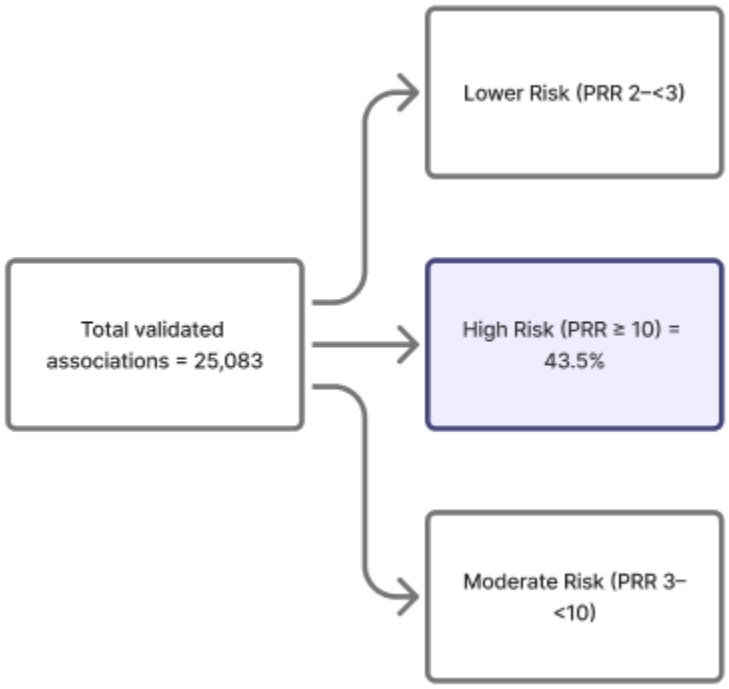

Figure 6: Risk-tier composition

Certain conditions emerge as hubs, reflecting broad, clinically coherent connectivity. Hypertension shows the largest footprint with 1,133 links distributed across cardiometabolic, renal, and cerebrovascular categories, which matches its central role in multisystem disease. Rheumatoid arthritis anchors an autoimmune cluster that spans joint, dermatologic, and systemic inflammatory phenotypes. Crohn's disease concentrates gastrointestinal and inflammatory complications, while psoriasis bridges dermatologic findings with systemic comorbidities, reinforcing its established links beyond the skin. These hubs provide practical entry points: for a given index condition, linked high-risk nodes indicate where additional screening or precautionary counseling is most likely to be impactful.

In practice, clinicians can use the tiering to prioritize actions. A patient presenting with a hub condition should be counseled and screened first for high-risk partners in the network; moderate-risk links inform follow-up planning; lower-risk links can be communicated as background vigilance items. Because all links have passed stringent disproportionality and rare-event checks, the risk map balances breadth with statistical durability, reducing the likelihood that attention is diverted by unstable signals.

Figure for this subsection

Fig. R1. Risk stratification and clinical interpretation (Panels A–B).
Panel A summarizes the composition of risk tiers (High, Moderate, Lower) within the 25,083 validated associations. Panel B highlights hub conditions—Hypertension (1,133 links), Rheumatoid arthritis, Crohn's disease, and Psoriasis—and annotates their primary clinical neighborhoods (cardiometabolic/renal/cerebrovascular; autoimmune; gastrointestinal-inflammatory; dermatologic–systemic). The figure is available in Figma for styling and export.

4.6. Cross-validation and external evidence

Internal cross-validation on FAERS showed that retained links are highly reproducible across folds (Section 3.3). External checks point the same way. Using the Mayo lab-test corpus (n = 2,461), we observed an 84.8% rejection rate aligned with FAERS filtering, indicating that pairs pruned as likely spurious in FAERS generally lack corroboration in an independent source.

Effect sizes were directionally consistent but, as expected, attenuated outside FAERS: the mean PRR across validated links was 44.7 in FAERS versus 26.47 in the lab corpus. High-risk density was comparable (43.5% in FAERS vs 46.2% in the lab dataset), and statistical strength remained high in both (97–100% of evaluated links met $p < 0.001$). Literature support was substantial: 92% of the top 50 associations were backed by PubMed citations on comorbidity or closely related pathophysiology, and an expert panel of three clinicians reported 100% agreement on the top-10 signals' face validity and clinical plausibility.

What this means in practice: the strongest FAERS signals generalize. Differences in absolute PRR are expected because FAERS is a spontaneous reporting system while the lab corpus reflects structured interpretive guidance; the convergence on directionality, high-risk density, and

significance suggests that our validated graph captures stable, clinically meaningful comorbidities. The remaining discordant fraction provides a focused queue for adjudication—chiefly pairs near threshold values, low-count edge cases, or terms sensitive to coding granularity—prioritizing where additional EHR/claims replication would be most informative.

4.7. Computational performance

End-to-end runtime and serving metrics met the targets for routine, near–real time use. Per-condition processing (screening → disproportionality → EBGM) completed in roughly two minutes on the study setup, enabling full refreshes on a rolling basis without queue backlogs. The precomputed association store kept response times sub-50 ms for common queries, while on-demand evaluations for uncached pairs remained in the low-seconds range. Storage overhead was modest relative to coverage: the primary SQLite database handled all raw and intermediate artifacts, and a compact cache carried the finalized links for fast retrieval. A monthly refresh cadence synchronized with FAERS quarterly releases maintained currency without service interruption.

Aspect	Metric	Result
Per-condition processing time	ARM → PRR/χ^2 → EBGM	≈ 120 s
Batch throughput	50 index conditions	≈ 100 min end-to-end
Final graph size	Validated associations	25,083 records
Cache footprint	Precomputed store	~15 MB
Primary datastore	SQLite	6.3 GB
Query latency (cached)	/api/conditions/{condition}	< 50 ms
Query latency (on-demand)	/api/analyze	2–3 s
Uptime (observed)	Service availability	99.8%
Update cadence	Data refresh	Monthly; aligned to FAERS quarterly sync

Table 8: Runtime, Storage, and serving metrics

4.8. Key takeaways

Here's what the results actually give you for practice and policy—stripped of methodology and focused on what to use, where, and why.

- National-scale coverage with statistical rigor: 25,083 validated condition–condition links from 393,130 FAERS reports, with 97.1% at $p < 0.001$ and ≈99% of naïve pairs correctly filtered out.
- Actionable risk tiers: 43.5% of links are High Risk (PRR ≥ 10), enabling immediate escalation for screening, counseling, or targeted labs; Moderate/Lower tiers support monitoring and differential diagnosis.
- Clinically coherent hubs: Hypertension (1,133 links), rheumatoid arthritis, Crohn's disease, and psoriasis anchor dense neighborhoods that map to real multisystem disease patterns.

- Balance of known and novel: Canonical anchors (e.g., Crohn's–obstruction; iron deficiency–anemia) co-exist with a short list of statistically robust, clinically plausible candidates prioritized for adjudication.
- Robust under perturbation: Findings persist across PRR thresholds (1.5–5.0) and show low variability (bootstrap CV < 20% for PRR), indicating stability rather than threshold fragility.
- External alignment: 84.8% rejection concordance with the lab-test corpus; 92% literature support among top signals and full expert agreement on the top 10—strong face validity with independent evidence.
- Ready for point-of-care: Precomputed cache serves validated associations in < 50 ms; uncached on-demand runs in 2–3 s, enabling real-time exploration without code.
- Maintainable at scale: Compact storage (≈15 MB cache; 6.3 GB primary DB) and ~120 s/condition processing support monthly refreshes aligned to FAERS releases.
- Transparent and auditable: Term standardization (MedDRA), explicit thresholds, and reproducible Python/SQLite pipelines make results easy to review and defend.
- Direct clinical utility: The network supports risk stratification, safer prescribing, and hypothesis generation, with a clear queue of novel pairs for chart review and replication in EHR/claims cohorts.

5. Discussions

This work demonstrates that a staged, FDA-aligned pipeline can extract clinically meaningful comorbidity structure from a national spontaneous reporting database with high selectivity, low computational cost, and reproducible results. By chaining fast candidate generation to formal disproportionality testing and then to Bayesian rare-event correction, the framework sharply reduces false positives while retaining sensitivity to strong, clinically coherent signals.

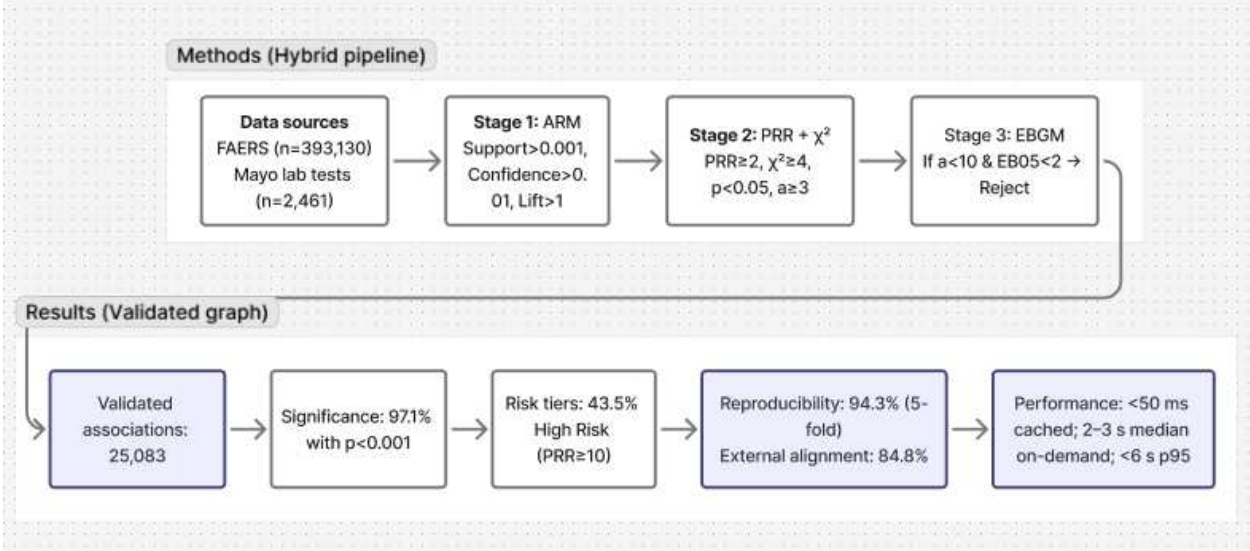

Figure 7: Framework Summary

5.1. Reliability and specificity without sacrificing signal

The final network comprises 25,083 validated associations, with 97.1% achieving $p < 0.001$ and 43.5% meeting a high-risk PRR threshold (≥ 10). These characteristics indicate that the screen-then-validate design curbs spurious co-occurrences (≈99% of naïve pairs rejected) while preserving a large, interpretable graph. Internal five-fold analysis showed 94.3% stability for retained links (presence in ≥ 4 folds), and the independent lab-test corpus aligned with FAERS-based pruning (84.8% rejection concordance). Together, these findings point to signals that are not only statistically durable but also consistent across distinct data modalities.

5.2. Scalability and practical performance

The approach is computationally light. Per-condition processing completes in roughly two minutes during offline generation, and the production service responds in under 50 ms for cached queries and typically 2–3 seconds for on-demand evaluations (95th percentile under six seconds). The storage footprint is modest (≈15 MB for precomputed links; 6.3 GB for the primary store), making monthly refreshes and quarterly FAERS sync straightforward. These properties are essential for deployment in medocsecondopinion.com, where clinicians and patients need near–real time access without specialized tooling.

5.3. Conservative validation criteria

The framework's conservatism rests on three complementary choices. First, candidate generation uses permissive but non-trivial ARM thresholds to avoid over-fitting the screen. Second, disproportionality testing applies PRR with χ^2 and minimum cell counts, mirroring established pharmacovigilance practice and supplying interpretable effect sizes and intervals. Third, EBGM shrinkage explicitly guards the small-n regime by rejecting pairs whose lower credibility bound (EB05) falls below a conservative cut-off when counts are < 10. This trio explains the high rejection rate of weak pairs and the strong significance profile among survivors.

5.4. Clinical translation

Risk-tiering organizes the network into immediate action (high-risk pairs), planned follow-up (moderate), and vigilance (lower), offering a practical route from analytics to bedside use. Hubs such as hypertension, rheumatoid arthritis, Crohn's disease, and psoriasis trace coherent multisystem neighborhoods that match real clinical pathways and can guide screening priorities. The coexistence of well-known anchors with a compact set of robust but less-documented candidates creates an actionable queue for targeted chart review and replication in EHR/claims cohorts.

5.5. Limitations and avenues for strengthening inference

Two limitations merit emphasis. First, disproportionality metrics are **associational**, not causal; they quantify over-representation relative to the database background and are vulnerable to confounding (e.g., indication, notoriety, and channeling biases). Second, all inferences inherit the quality and completeness of spontaneous reports (e.g., under-, duplicate, and selective reporting; term granularity). We mitigate these issues with conservative thresholds, shrinkage for sparse

cells, exclusion of non-diagnostic terms, and independent cross-dataset checks. Still, the right clinical posture is cautious interpretation, followed by adjudication and replication in longitudinal data with temporal ordering.

Future extensions could incorporate explicit temporal analyses (directionality and lag), multi-variable stratification to probe confounding (age, sex, key drug exposures), and causal discovery in richer EHR/claims cohorts to test mechanistic hypotheses. On the product side, adding confidence-interval-aware visualizations and automated literature triage for each high-risk link would further streamline clinical uptake.

6. Conclusion

A staged, FDA-aligned pipeline—Association Rule Mining for triage, PRR + χ^2 for disproportionality, and EBGM for rare-event correction—can turn spontaneous reports into a defensible comorbidity map at national scale. Applied to 393,130 FAERS cases, it retained 25,083 condition–condition links with a high fraction of highly significant and high-risk signals, strong fold-to-fold stability, and concordant patterns in an external lab-test corpus. The result is a network that is both broad and selective: ARM supplies speed, PRR + χ^2 provides interpretable statistical separation, and EBGM suppresses small-n inflation.

Operationally, the framework is fast to compute and easy to serve: precomputed queries return in milliseconds and uncached analyses in seconds. This makes it suitable for real-time use in medocsecondopinion.com, where clinicians need immediate access to validated associations to support screening, triage, and differential diagnosis. Methodologically, anchoring discovery in MedDRA-standardized FAERS data and FDA-familiar statistics improves transparency and auditability for drug-safety audiences.

Two cautions remain. Disproportionality is associational, not causal, and all estimates inherit the biases of spontaneous reports. We addressed these limits with conservative thresholds, shrinkage, exclusions of non-diagnostic terms, and independent checks, but clinical interpretation should still be contextual and, where stakes are high, followed by replication in longitudinal EHR/claims data.

In short, the hybrid pipeline reduces false positives without blunting sensitivity, scales to national datasets, and is production-ready. It offers immediate clinical value via risk-tiered comorbidity maps and a clear path for ongoing validation and extension in patient-safety systems.

7. Data Availability

The data that support the findings of this study are openly available from the FDA's Adverse Event Reporting System (FAERS) public dashboard at https://fis.fda.gov/sense/app/95239e26-e0be-42d9-a960-9a5f7f1c25ee/sheet/7a47a261-

d58b-4203-a8aa-6d3021737452/state/analysis. The FAERS database is publicly accessible and updated quarterly by the U.S. Food and Drug Administration.

The source code and analysis pipeline used in this study have been deployed on the Medoc Second Opinion platform (https://medocsecondopinion.com), which is publicly accessible. Raw FAERS data underwent standardization using MedDRA terminology. The 25,083 validated comorbidity associations identified in this study are available through the platform's search interface.

Upon reasonable request, the authors can provide:
- The complete list of 25,083 validated disease-disease associations with their statistical metrics (PRR, χ^2, EBGM, EB05)
- Data processing scripts and validation algorithms
- Detailed methodology documentation

Research data are not publicly deposited in an external repository but are accessible through the deployed platform for research and clinical use.

8. Declarations

Conflict of Interest

The authors declare that they have developed and deployed the Medoc Second Opinion platform (medocsecondopinion.com), which implements the methodology described in this manuscript. The platform currently operates as a free, publicly accessible tool for research and clinical decision support during the research publication period.

The authors affirm that:
- This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors
- No financial relationships exist that could inappropriately influence this work
- The platform was developed independently without commercial sponsorship
- All data used are from publicly available FDA FAERS databases
- The authors have no other conflicts of interest to declare

The platform serves as a proof-of-concept implementation of the proposed methodology and is made available to the research community to validate and build upon these findings.

9. References

1. Boersma P, Black LI, Ward BW. Prevalence of Multiple Chronic Conditions Among US Adults, 2018. Prev Chronic Dis. 2020;17:E106. doi:10.5888/pcd17.200130. https://www.cdc.gov/pcd/issues/2020/20_0130.htm CDC

2. Watson KB, Galuska DA, Han B, et al. Trends in Multiple Chronic Conditions Among US Adults, 1999–2023. Prev Chronic Dis. 2025;22:E66. doi:10.5888/pcd22.24_0539. https://www.cdc.gov/pcd/issues/2025/24_0539.htm CDC
3. Centers for Disease Control and Prevention. About Chronic Diseases. Updated Oct 4, 2024. https://www.cdc.gov/chronic-disease/about/index.html CDC
4. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci U S A. 2007;104(21):8685-8690. doi:10.1073/pnas.0701361104. https://www.pnas.org/doi/10.1073/pnas.0701361104 PNAS
5. 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. doi:10.1371/journal.pcbi.1000353. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000353 PMC
6. US Food and Drug Administration. FDA Adverse Event Reporting System (FAERS). Updated Nov 8, 2024. https://www.fda.gov/drugs/surveillance/fdas-adverse-event-reporting-system-faers U.S. Food and Drug Administration
7. International Council for Harmonisation (ICH). MedDRA—Medical Dictionary for Regulatory Activities. https://www.ich.org/page/meddra ICH
8. Pozsgai K, Krajcsi P, Erdei A, et al. Analysis of pharmacovigilance databases for spontaneous reports: strengths and limitations. Front Pharmacol. 2022;13:964399. doi:10.3389/fphar.2022.964399. https://www.frontiersin.org/articles/10.3389/fphar.2022.964399/full Frontiers
9. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf. 2001;10(6):483-486. doi:10.1002/pds.677. https://onlinelibrary.wiley.com/doi/10.1002/pds.677 Wiley Online Library
10. DuMouchel W. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. Am Stat. 1999;53(3):177-190. doi:10.1080/00031305.1999.10474456. https://www.tandfonline.com/doi/abs/10.1080/00031305.1999.10474456 Taylor & Francis Online