

Safety and Adverse Event Characterization of Type 2 Diabetes Treatments

Final Project Write-up

Team Members: Jiayi Chu, Yixin Xue, Shuya Guo, Runjiu Chen

2025-12-16

Overview

Project Title: Safety and Adverse Event Characterization of Type 2 Diabetes Treatment

Team Name: Jet2Holiday

Team Members:

- Jiayi Chu - jchu47
- Yixin Xue - yxue39
- Shuya Guo - sguo66
- Runjiu Chen - rchen137

Background

Type 2 diabetes mellitus (T2DM) is a chronic disease that typically requires long-term pharmacologic management, often involving the concurrent or sequential use of multiple medications. Current treatment options span several mechanistic classes. Because these drug classes act through distinct biological pathways, they are associated with heterogeneous adverse event (AE) profiles, ranging from gastrointestinal and metabolic effects to cardiovascular, renal, and neurological outcomes. As newer therapies are increasingly prescribed alongside established agents, the overall safety landscape of T2DM treatment has become progressively more complex.

In this context, real-world pharmacovigilance data play an essential role in complementing evidence from randomized clinical trials. The FDA Adverse Event Reporting System (FAERS) is one of the largest publicly available spontaneous reporting databases and has been widely used for post-marketing safety surveillance across therapeutic areas. FAERS captures adverse events reported by patients, healthcare providers, and manufacturers and encodes clinical outcomes using the Medical Dictionary for Regulatory Activities (MedDRA), enabling large-scale assessment of drug safety in routine clinical practice. For chronic diseases such as diabetes, FAERS provides a valuable opportunity to evaluate rare, delayed, or population-specific adverse events that may not be fully characterized in pre-approval trials.

A growing body of literature has applied FAERS data to investigate the safety of T2DM medications, including analyses of individual drugs and specific mechanistic classes. These studies have

identified potential safety signals related to gastrointestinal, cardiovascular, renal, and ophthalmic outcomes and have demonstrated the feasibility of using disproportionality-based methods for signal detection in diabetes pharmacotherapy^{1,2}. However, the existing literature remains fragmented. Most analyses focus on a single drug or a narrowly defined adverse event, limiting the ability to systematically compare safety patterns across therapeutic mechanisms.

In addition, FAERS data present substantial analytic challenges that complicate cross-drug interpretation. Drug names are inconsistently reported, adverse events are distributed across hierarchical MedDRA levels, and reports are stored across multiple relational tables. Without standardized preprocessing pipelines and unified analytic frameworks, results from FAERS-based studies can be difficult to reproduce, compare, or translate into clinically meaningful insights.

These limitations are particularly important given that diabetes medications are often used chronically and in combination, and even modest differences in adverse event profiles may have significant implications at the population level. Together, this body of work highlights the need for a reproducible, scalable approach that integrates FAERS data across drug classes and presents adverse event patterns in a transparent and interpretable manner.

References

1. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). *Expert Opin Drug Saf.* 2017;16(7):761–767. doi:10.1080/14740338.2017.1323867.
2. Cheng X, Jiang Z, Li G, Wang J, Han F. Multi-database pharmacovigilance assessment of GLP-1 receptor agonist-related ophthalmic risks using advanced signal detection in FAERS and VigiBase. *Journal of Endocrinological Investigation.* 2025; Published online September 29, 2025:1–9. doi:10.1007/S40618-025-02712-3.

Orginal Goals

The goal of this project was to address these challenges by developing a scalable and reproducible R-based analytic workflow to characterize real-world adverse event patterns associated with Type 2 diabetes medications using FAERS data from 2019 to 2021. Specifically, we aimed to:

- Build an automated data-processing pipeline that integrates and cleans multi-year FAERS data
- Standardize drug names and MedDRA-coded adverse events to enable consistent cross-drug and cross-class comparisons
- Quantify and compare adverse event patterns at both the mechanistic class level and individual drug level
- Identify potential safety signals and temporal patterns using descriptive summaries and disproportionality-based approaches
- Deliver results through an interactive Shiny dashboard that supports transparent exploration and monitoring of T2DM drug safety

Existing Work

Post-marketing safety studies using the FDA Adverse Event Reporting System (FAERS) have been widely used to evaluate the safety profiles of medications for type 2 diabetes mellitus. FAERS provides information on patient demographics, drug exposures, reporting dates, and MedDRA-coded adverse events, enabling descriptive summaries, disproportionality analyses, and assessments of temporal trends. Previous studies have established FAERS as an appropriate data source for pharmacovigilance research in this therapeutic area.

Most existing FAERS-based studies focus on individual drugs or a single drug class, particularly glucagon-like peptide 1 receptor agonists. Several recent papers have examined associations between GLP 1 receptor agonists and specific outcomes such as ocular adverse events, psychiatric symptoms, and acute kidney injury. While these studies provide detailed evaluations of targeted safety concerns, their narrow scope limits cross-class comparisons across major type 2 diabetes treatments.

In addition, many prior analyses rely on manual or semi-manual data processing workflows, which can reduce reproducibility and limit scalability. Comparative analyses that simultaneously evaluate multiple type 2 diabetes drug classes, especially at both the drug and mechanistic class levels, remain relatively uncommon. Temporal trend analyses and forecasting approaches are also rarely incorporated. Our project addresses these gaps by conducting a unified, automated analysis across all major type 2 diabetes drug classes using a reproducible R-based workflow and an interactive Shiny dashboard.

Data

Data for this study were obtained from the FDA Adverse Event Reporting System (FAERS), a publicly available spontaneous reporting database that collects post-marketing safety reports for approved drugs and therapeutic biologics. We used quarterly FAERS ASCII datasets covering the period from 2019 to 2021. These datasets include multiple relational tables describing patient demographics, drug exposures, reported adverse events, and clinical outcomes. All quarterly files were programmatically downloaded and extracted using R, and organized by year and quarter to ensure a reproducible data acquisition process.

After obtaining the raw FAERS data, we implemented a systematic data cleaning and standardization workflow. Column names were standardized across all tables to enable consistent downstream integration. Field contents were cleaned to improve data quality, including trimming extraneous whitespace, normalizing text case, parsing date fields into consistent formats, and applying reasonable bounds to numeric variables. To address duplicate reporting, multiple versions of the same case were consolidated by retaining only the most recent report, ensuring that each case contributed a single record to the analysis.

All cleaned tables were then imported into a relational SQLite database using the DBI and RSQLite packages. Storing the data in a database allowed efficient handling of the large-scale FAERS dataset and supported complex joins and aggregations required for analysis. Additional validation steps were performed at the database level to remove implausible records, such as biologically unrealistic age values or inconsistent event timelines, in order to improve the stability and reliability of downstream analyses.

Target type 2 diabetes mellitus (T2DM) medications were identified using a keyword-based wide screening approach. Because drug names in FAERS are free-text and highly heterogeneous, we standardized drug names using the RxNorm API, with local caching to improve efficiency and reproducibility. The analytic cohort was restricted to reports in which one of the selected drugs was designated as the primary suspect, resulting in a final set of 15 target T2DM medications.

Adverse events in FAERS are reported using MedDRA Preferred Terms (PTs), which are highly granular and difficult to compare directly across drugs. To facilitate organ-level comparisons and visualization, PTs were mapped to a pre-defined system organ–level classification constructed from publicly available resources. This approach enabled meaningful aggregation of adverse events while remaining transparent and reproducible.

Finally, patient demographics, drug exposures, adverse events, and outcome indicators were integrated into a case-level analytic dataset. Drug exposures were further annotated with WHO ATC codes and mechanistic class labels, such as GLP-1 receptor agonists and SGLT2 inhibitors. The resulting analysis-ready table served as the unified data source for all downstream statistical analyses and dashboard visualizations in this study.

Study Design

We conducted a retrospective observational pharmacovigilance study using data from the FDA Adverse Event Reporting System (FAERS). FAERS is a spontaneous reporting database that collects adverse event reports submitted by healthcare professionals, manufacturers, and consumers, and includes information on patient demographics, drug exposures, reported adverse events, outcomes, and reporting dates. For this study, we analyzed quarterly FAERS datasets covering the period from 2019 to 2021.

The study population consisted of FAERS reports in which one of the selected type 2 diabetes mellitus (T2DM) medications was listed as the primary suspect drug. Exposure was defined at both the individual drug level and the mechanistic class level, with drugs grouped according to Anatomical Therapeutic Chemical (ATC) classifications. Adverse events were defined as outcomes reported in FAERS and coded using MedDRA Preferred Terms, which were further summarized into pre-defined System Organ Class categories to facilitate organ-level comparisons. The primary analytic focus was on characterizing and comparing adverse event reporting patterns across drug classes and individual drugs rather than estimating causal effects. Analyses included descriptive summaries, disproportionality measures such as reporting odds ratios and proportional reporting ratios, and assessments of temporal reporting trends to explore potential safety signals in real-world data.

Technical Challenges

Large database size: Processing multi-million FAERS records required database-backed cleaning to overcome memory constraints.

Drug name normalization via external APIs: Integrating the RxNorm API to standardize heterogeneous drug names while handling rate limits, network latency, and caching for reproducibility.

Adverse event classification without licensed MedDRA SOC: Due to lack of access to the official MedDRA PT–SOC mapping, we constructed a custom, literature-informed SOC mapping to

enable organ-level analysis.

Programming Paradigms

Paradigm	Where used	Purpose
Command line	R script step 1 to step 11	Batch ETL and cleaning
Functional (dplyr / purrr)	Pipeline scripts, cohort assembly	Composable data transforms
Database (DBI + SQLite)	Cohort analytic and staging tables	Large joins and aggregation
Machine learning / Time series	Page 5: ARIMA, XGBoost, RF, ENet	Emerging signal forecasts

Shiny APP Overview

Page 1 Home: Introduces the study background, data source, target drugs and mechanisms, research questions, and the overall FAERS based analytic workflow used in the dashboard.

Page 2 Global Trends: Presents overall adverse event reporting patterns from 2019 to 2021, including reporting volume, temporal trends, seriousness distributions, and comparisons across major drug classes.

Page 3 Mechanism Comparison: Compares adverse event patterns across drug classes using SOC distributions, disproportionality metrics, and rankings of commonly reported preferred terms.

Page 4 Individual Drug Profiles: Provides drug specific safety summaries for each included medication, showing reporting trends, frequently reported adverse events, SOC distributions, and signal detection results.

Page 5 Temporal & Emerging Signals: Explores longitudinal reporting patterns and potential emerging safety signals using machine learning based time series models.

Page 6 Methods & Downloads: Describes dashboard ownership, integrated programming paradigms, data scope, and key methodological limitations of using FAERS for pharmacovigilance analysis.

Usability, Documentation, and Reproducibility

The Shiny dashboard was designed to support clear and intuitive exploration of adverse event patterns across type 2 diabetes treatments. The dashboard follows a page-based structure, with each page focusing on a specific analytical task, including global trends, mechanistic class comparisons, individual drug profiles, and temporal signal monitoring. Interactive elements such as dropdown menus and filters allow users to select drugs, classes, and time periods of interest, with outputs updating reactively to ensure consistency across visualizations.

To improve usability and interpretability, the dashboard includes descriptive text, clear axis labels, and sufficiently large figure titles and legends. Key analytic measures, including adverse event counts

and disproportionality metrics, are explained within the interface to support correct interpretation by users who may not have prior experience with pharmacovigilance data. A dedicated methods page summarizes the data processing pipeline, drug classification strategy, and analytical approach used in the project.

All code and materials are organized in a GitHub repository submitted through GitHub Classroom. The repository includes a README.md file that summarizes the project objectives, provides an overview of the analytical workflow, and lists team members and their contributions. Instructions for running the data processing pipeline and launching the Shiny dashboard locally are also provided. Source code is structured into clearly labeled scripts corresponding to data acquisition, preprocessing, analysis, and dashboard development.

Reproducibility was emphasized throughout the project. All analyses are implemented through scripted workflows, and links to the original FAERS data source are included in the repository. By combining automated pipelines, clear documentation, and accessible code, the project supports reproducible and transparent analysis of real-world adverse event data.

Results and Key Findings

Global Trend

The Global Trends module provides a descriptive overview of AE reporting patterns across time, severity, mechanism class, and clinical domains.

Overall, quarterly reporting volumes remain high and relatively stable over the study period. Total AE counts fluctuate between approximately 17,000 and 21,000 reports per quarter, with noticeable short-term variation but no sustained decline. A marked dip appears in early 2020, followed by a recovery in subsequent quarters, indicating persistent real-world reporting activity for T2DM therapies.

When stratified by seriousness, non-serious cases consistently outnumber serious cases across all quarters. Non-serious reports generally range from about 11,000 to 13,000 per quarter, while serious reports remain lower, around 6,500 to 8,500. Despite the lower absolute counts, serious events constitute a substantial and stable proportion of total reports over time, underscoring the clinical relevance of the dataset beyond mild outcomes.

Event counts differ markedly across mechanism classes. Biguanides account for the largest share of reported events, followed by DPP-4 inhibitors and basal insulin. This pattern highlights substantial heterogeneity in reporting burden across drug classes and motivates mechanism-specific safety monitoring.

Finally, analysis of the top-reported pseudo System Organ Classes (SOCs) shows that a small number of clinical domains dominate AE reporting across all drugs. Immune system disorders, gastrointestinal disorders, general disorders, and nervous system disorders appear most frequently, followed by metabolic, vascular, and renal-related categories. These SOCs define the primary clinical areas in which adverse events are most commonly reported and provide a high-level map for prioritizing more detailed downstream analyses.

Mechanism Comparison

Clear differences in adverse event reporting patterns were observed across major type 2 diabetes drug classes. Heatmap visualizations of preferred terms, scaled within each mechanism, showed

that the distribution of reported adverse events varied across drug classes. Within each mechanism, a small number of preferred terms accounted for a larger share of reports, while many other events were reported less frequently. However, the specific adverse events that appeared most prominently differed across mechanisms, indicating that reporting patterns were not uniform across drug classes.

Examination of the most frequently reported adverse events within each mechanistic class further highlighted these differences. Several preferred terms, such as nausea, fatigue, and dizziness, appeared among the top reported events across multiple drug classes. At the same time, other adverse events were more specific to individual mechanisms. For example, injection site related events were more prominent among glucagon like peptide 1 receptor agonists, while other classes showed different sets of frequently reported events. These results reflect heterogeneity in within class reporting profiles across mechanisms.

Disproportionality analyses using proportional reporting ratios and reporting odds ratios provided additional insight into these patterns. For certain preferred terms, disproportionality metrics were higher within specific mechanistic classes compared with others. The magnitude of these metrics varied across preferred terms and drug classes. These analyses were used to identify potential safety signals and to complement descriptive summaries rather than to establish causal relationships.

Individual Drug Profiles

Across the 15 type 2 diabetes medications analyzed, individual drug profiles revealed substantial heterogeneity in reporting volume, temporal patterns, adverse event composition, and severity. Some drugs exhibited relatively stable or declining reporting trends over the study period, while others showed pronounced fluctuations or late-period increases. The proportion of serious adverse events varied widely across drugs, ranging from approximately 14% to over 55%, indicating marked differences in reported severity even among treatments for the same condition.

In terms of reporting volume and temporal trends, long-established drugs with broad patient exposure, such as metformin and insulin glargine, accounted for the largest number of reports between 2019 and 2021, with total case counts reaching the tens of thousands. In contrast, newer or less frequently prescribed agents, including ertugliflozin and saxagliptin, generated substantially fewer reports. Several GLP-1 receptor agonists, notably semaglutide and dulaglutide, demonstrated clear increases or peaks in reporting during later quarters, consistent with expanding clinical adoption, whereas older agents such as exenatide and some DPP-4 inhibitors showed gradually declining trends.

Adverse event composition at the system organ class (SOC) level was broadly consistent across drugs but exhibited mechanism-specific patterns. For most medications, reports were dominated by immune system disorders, gastrointestinal disorders, and general disorders. GLP-1 receptor agonists showed a prominent gastrointestinal signal, aligning with known tolerability profiles, while SGLT2 inhibitors exhibited relatively higher proportions of infection-related and metabolic adverse events. DPP-4 inhibitors tended to display narrower SOC distributions, suggesting more concentrated adverse event profiles. These SOC-level patterns were stable across time scales.

Marked differences were also observed in serious adverse events. Certain DPP-4 inhibitors, such as alogliptin and saxagliptin, showed the highest proportions of serious cases, exceeding 50%, whereas GLP-1 receptor agonists like dulaglutide and semaglutide exhibited comparatively lower serious event proportions. Across all drugs, hospitalization was the most frequently reported serious outcome, followed by life-threatening events and death, while congenital anomalies were rare. Importantly, drugs with lower overall reporting volume did not necessarily have lower seriousness

proportions, highlighting the importance of considering both absolute counts and relative severity.

Taken together, individual drug-level analyses demonstrate that T2DM medications differ not only in how frequently adverse events are reported, but also in the types and severity of those events in real-world data. These findings underscore the value of drug-specific pharmacovigilance analyses and illustrate how the interactive framework developed in this study facilitates transparent and systematic comparison of safety profiles across medications.

Temporal & Emerging Signals

The Temporal & Emerging Signals module extends the descriptive and disproportionality analyses by adding a longitudinal view of reporting behavior. For each selected high-volume drug, we aggregate FAERS reports by calendar quarter (2019–2021) and visualize the observed time series of report counts. This component is designed to help users identify (i) gradual changes in reporting over time, (ii) short-term spikes or dips, and (iii) whether a drug's recent pattern appears unusual relative to model-based expectations.

To enable model-based comparison, the dashboard implements a multi-model forecasting overlay that predicts reporting counts for the next four quarters and displays results both as an overlaid plot and a numeric table. We include multiple forecasting paradigms to reflect different assumptions about trend smoothness and sensitivity to irregular fluctuations:

- **ARIMA (statistical baseline):** captures autocorrelation structure and smooth temporal dependence, serving as a classical reference model.
- **XGBoost (machine learning):** flexible nonlinear model that can adapt to irregular quarter-to-quarter variation.
- **Random Forest (ensemble learning):** robust nonlinear baseline that tends to be stable under noisy patterns.
- **Elastic Net (regularized regression):** interpretable baseline emphasizing stable trends and reducing overfitting.

For visualization robustness, models that produce degenerate “flat-line” forecasts (nearly constant predictions across all future quarters) are automatically filtered out, since such outputs are often uninformative for monitoring and can dominate the plot visually. When one or more models fail, the module continues to render the observed series and any available forecasts, ensuring stability of the user interface.

Across the top drugs, forecasts varied substantially by model, reflecting different inductive biases. In general, machine learning models (especially XGBoost) tended to be more responsive to nonlinear changes and short-term variation, while Elastic Net produced smoother projections and Random Forest provided moderate flexibility with stable behavior. ARIMA served as a conservative statistical baseline, which can be appropriate for stable series but may under-adapt when trends change rapidly. This multi-model view supports exploratory interpretation: when models agree, the projected trend appears more robust; when models diverge, it suggests higher uncertainty or sensitivity to recent fluctuations.

Importantly, this module is intended for exploratory signal monitoring rather than causal inference. FAERS reporting volumes are influenced by utilization, publicity, reporting practices, and other external factors. Therefore, forecast deviations should be interpreted as prompts for further investigation (e.g., stratification by event category, seriousness, or mechanistic class), not as definitive

evidence of increased drug risk.

Limitations and Future Work

Limitations

- FAERS is a spontaneous reporting system and is subject to underreporting and reporting bias. Reported adverse events do not represent true incidence and cannot be used to estimate risk.
- Drug and adverse event information may be misclassified, and adverse events were summarized using a custom MedDRA mapping rather than licensed SOC classifications.
- Disproportionality measures and time series analyses are descriptive tools for signal detection and do not support causal inference.

Future Work

- Future work could integrate external data sources, such as prescription or claims data, to better contextualize reporting patterns.
- Access to official MedDRA mappings and longer observation periods could improve adverse event classification and temporal signal detection.

More for Our Product

- **Deployed Shiny App:** <http://4.155.111.178:3838>
- **Classroom GitHub Repository:** <https://github.com/jhu-statprogramming-fall-2025/project04-jet-2-holiday>
- **Personal GitHub Repository:** <https://github.com/Yixin-Xue/biostat777-Final-Project>

Some plots from the demo to show our app function

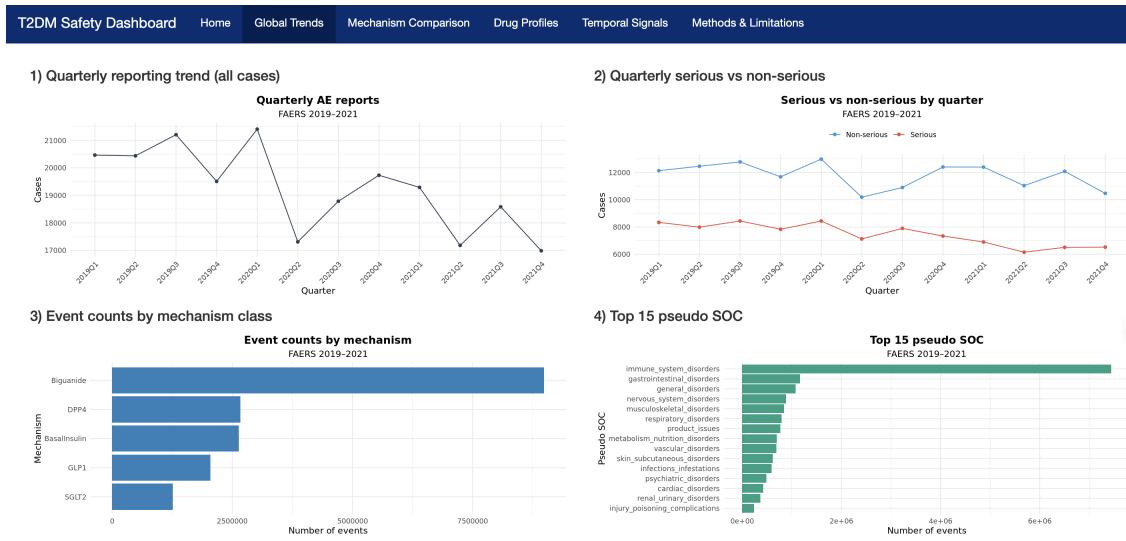


Figure 1: Global adverse event reporting trends for selected type 2 diabetes medications from 2019 to 2021.



Figure 2: Heatmap of the most frequently reported MedDRA preferred terms across major type 2 diabetes drug mechanisms, scaled within each mechanism. Users can adjust the number of displayed preferred terms, with the heatmap updating dynamically.

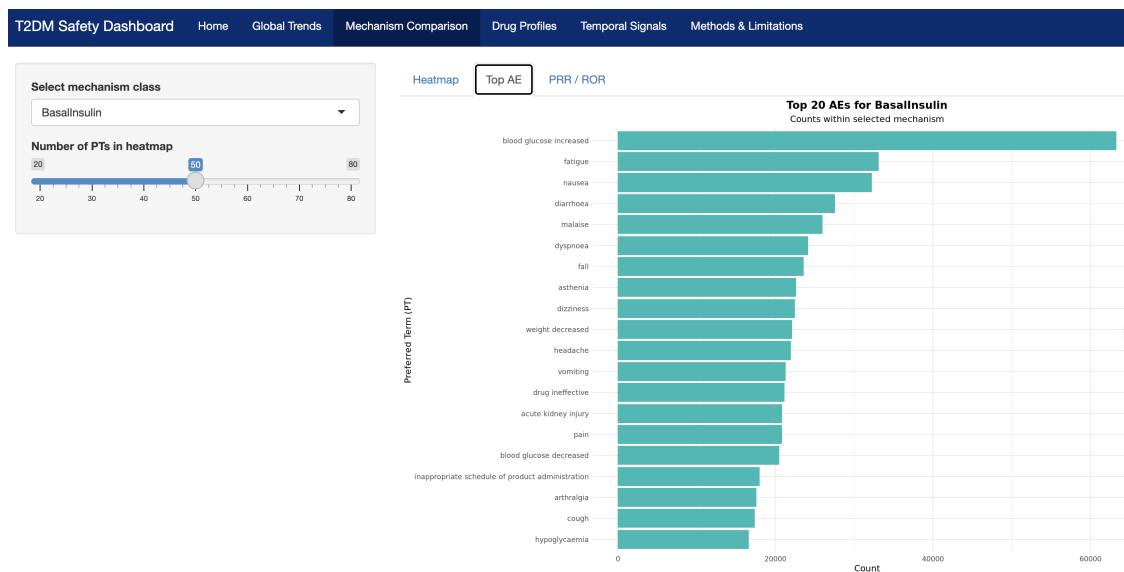


Figure 3: Top reported adverse events within each drug mechanism. Users can select a mechanism class, and the ranking of adverse events updates automatically based on the selected class.

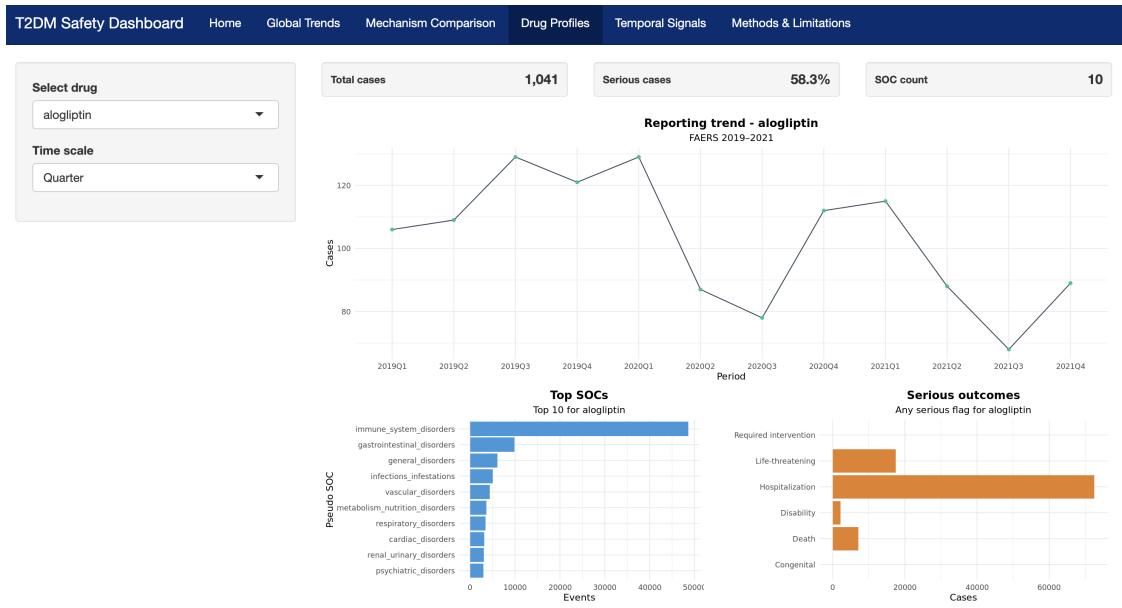


Figure 4: Individual drug safety profiles showing reporting trends, top adverse events, and adverse event distributions by system organ class. Users can select different drugs, and all tables and plots update reactively.

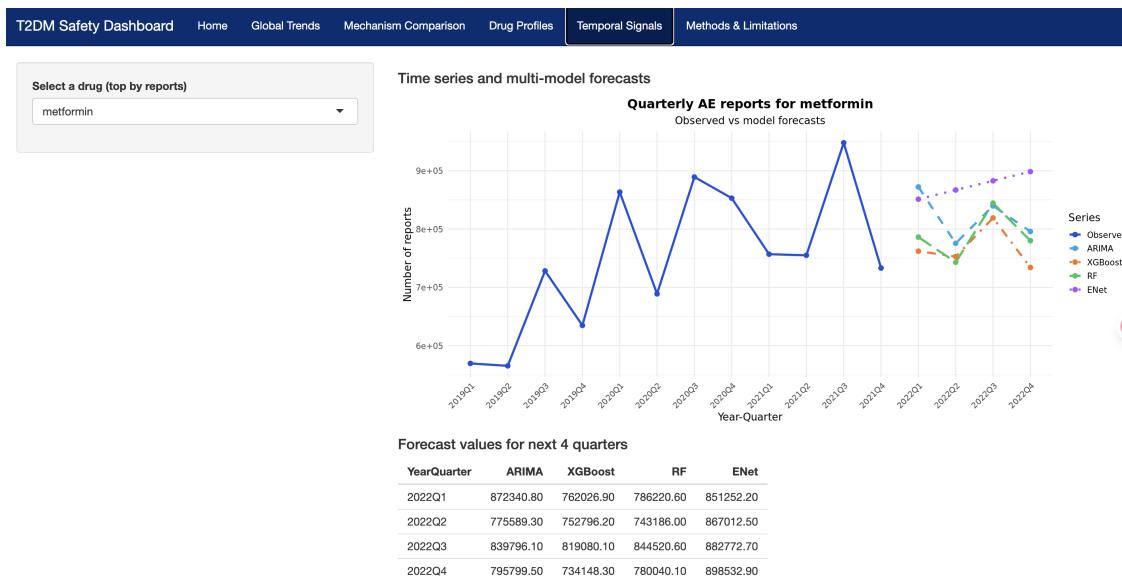


Figure 5: Temporal reporting patterns and machine learning based forecasts for selected drugs or mechanism classes. Time series models are used to identify potential emerging safety signals, and users can adjust selections to view updated trends and forecasts.