Sex-Specific Prediction of Adverse Drug Reactions Associated with Drug Combinations: A FAERS-Based Study

Xurui Gao
Brown University, May 2026

Master's Thesis

Submitted in fulfillment of the requirements for ScM in Biostatistics in the Brown University School of Public Health.

This thesis by Xurui Gao is accepted in its present form by the Brown University Department of Biostatistics as satisfying the thesis requirements for the degree of Sc.M. in Biostatistics.

Date _____ _____

Alice Paul, Advisor

Date _____ _____

Rebecca Hubbard, Reader

Date _____ _____

Stavroula Chrysanthopoulou, Director of the Master's Program in Biostatistics

Acknowledgements

I am deeply grateful to my advisor, Professor Alice Paul, for her invaluable guidance, patience, and continuous support throughout this thesis. Her expertise and insights have been instrumental in shaping both the direction and development of this work, and I have learned a great deal from working under her mentorship. I would also like to sincerely thank my reader, Professor Rebecca Hubbard, for her thoughtful comments and constructive feedback, which have significantly strengthened this research.

Abstract

Adverse drug reactions (ADRs) pose a significant threat to medication safety, with risks varying across populations. Sex is a key factor influencing the heterogeneity of drug responses. However, existing pharmacovigilance research often focuses on risk prediction in the overall population, lacking systematic modeling and comparative analysis of sex-specific effects. This study developed sex-specific ADR prediction models among individuals with reported ADR cases. The goal is to identify risk factors for specific ADRs and quantify differences between males and females, thereby improving the interpretability of pharmacovigilance and providing clinically meaningful sex-specific safety insights for personalized risk prediction.

Using the FAERS database (2015-2024), disproportionality analysis was employed to identify drugs significantly associated with each ADR. Two modeling strategies were subsequently applied for sex-stratified analysis: (i) Relaxed LASSO regression incorporating sex-drug interaction terms to estimate sex-specific effects within a linear framework; and (ii) sex-stratified Random Forest (RF) models combined with SHAP for interpretability, quantifying the predictive contribution of each drug in male and female models. Model performance was comprehensively evaluated using accuracy, ROC-AUC, PR-AUC, and F1 score.

The two modeling approaches exhibited distinct predictive performance and risk identification characteristics across sex subgroups. The Relaxed LASSO model achieved superior overall predictive performance, capturing sex-drug interactions and identifying sex-biased drugs (e.g., Striant in males, Zolpidem in females). Conversely, the RF model revealed nonlinear structures, identifying cross-sex core drugs and sex-specific contributors (e.g., follitropin in females, influenza virus in males). These two methods are complementary in capturing sex heterogeneity.

This study systematically revealed both shared and divergent ADR risk profiles across sexes via two interpretable models. Relaxed LASSO is well-suited for identifying and quantifying sex-drug interactions, while RF excels in uncovering complex risk structures. This multi-method comparison not only validates the necessity of sex-stratified modeling but also provides a methodological foundation for developing interpretable, robust, and clinically actionable sex-differentiated pharmacovigilance systems. The two methods identified distinct sets of risk-associated drugs, warranting further investigation. Future research should further integrate physiological, genomic, and other multidimensional data to advance the development of precision medication safety monitoring.

Contents

1 Introduction **1**

2 Data Description **3**

3 Data Preprocessing **4**

3.1 Data Overview 4

3.2 Data Cleaning and Filtering 6

3.3 Feature and Sample Selection 7

4 Methods **9**

4.1 Relaxed LASSO with Interaction Terms 9

4.2 Random Forest 12

5 Results **13**

5.1 Sex-Subgroup Model Performance 14

5.2 Sex-Specific Key Risk Drugs 15

5.3 Cross-Model Consistency of Sex-Specific Signals 23

6 Discussion **27**

6.1 Overall Findings: Method Dependence in Model Performance and Risk Identification 27

6.2 Sex-Differentiated Risk Profiles from a Methodological Perspective 28

6.3 Limitations 29

6.4 Future Directions 29

7 Conclusion **30**

A Appendix **35**

A.1 Sensitivity Analysis of the Impact of PRR Threshold Variation on Model Prediction Robustness 35

A.1.1 Research Purpose 35

A.1.2 Methodology 35

A.1.3 Results Analysis 36

A.2 Comparative Experiment on Modeling Strategies 37

1 Introduction

Adverse drug reactions (ADRs) are one of the leading causes of increased morbidity and mortality in clinical care, imposing substantial burdens on medication safety and public health [1, 2, 3, 4]. As the complexity of medication regimens increases, traditional pharmacovigilance systems face significant bottlenecks in the timely identification of individualized risks.

Importantly, ADR risk is not homogeneous, but depends on both drug properties and patient-specific factors. Among these factors, sex—through differences in anatomy, hormone profiles, and drug-metabolizing enzymes—has been repeatedly recognized as a key source of heterogeneity in drug responses. For example, [5] investigated sex-related patterns in contrast-media adverse reactions and reported distinct risk structures across sexes. [6] synthesized evidence on sex differences in ADRs and emphasized the importance of incorporating sex considerations throughout the drug development and surveillance lifecycle. In the context of COVID-19 therapeutics, [7] likewise observed significant sex differences in reported adverse events, underscoring the need to clarify the mechanistic and clinical roles of sex in ADR risk. Collectively, these findings motivate systematic efforts to characterize and quantify individual-level heterogeneity—and especially sex heterogeneity—in ADR risk as a core step toward advancing precision medicine and enhancing individualized medication safety. This study developed sex-specific ADR prediction models targeting users who have already submitted ADR reports. The goal is to identify risk factors for specific ADRs and quantify differences between males and females.

To mine drug safety signals from large-scale real-world data, the FDA Adverse Event Reporting System (FAERS) [8], one of the largest spontaneous reporting databases worldwide, has become a foundational resource for methodological advances in pharmacovigilance. In recent years, FAERS-based research has shifted from descriptive frequency-based analyses toward more sophisticated machine learning models, with increasingly fine-grained objectives that further push the frontier of drug–ADR surveillance methodology. A representative example is the work of [9], who constructed HODDI, a public dataset derived from a decade of FAERS reports and designed specifically to study higher-order drug interactions (i.e., complex interactions among multiple drugs). Their systematic benchmarking—ranging from multilayer perceptrons to hypergraph neural networks—demonstrated the feasibility of extracting informative features from the complex association structures of FAERS, providing a vital methodological reference for subsequent studies.

Other research has further moved toward refined clinical risk assessment. [10] integrated drug semantic information (e.g., indications and active substances) by mapping it into embedding representations using an enhanced BioWordVec framework, and combined these representations with machine learning to predict the severity of interaction-related adverse outcomes (serious vs. non-serious) directly from FAERS reports. A key contribution of this line of work is the transition from a binary "associated vs. not associated" paradigm to severity stratification with clearer clinical implications. However, although demographic variables such as age and sex were included as features,

their reported incremental benefit to overall predictive performance was limited. This suggests that simply adding demographics as generic covariates may be insufficient to capture subgroup-specific risk mechanisms in complex models.

Evidence for meaningful subgroup heterogeneity also extends beyond the FAERS database. For instance, [11] conducted a prospective multi-center study highlighting pharmacogenomic influences on ADR risk, while [12] leveraged the global VigiBase database to perform subgroup disproportionality analyses across regions and populations. Together, these studies reinforce the clinical value and necessity of subgroup-differentiated ADR risk assessment.

Despite these advances, most FAERS-based predictive modeling studies still primarily target population-level performance, aiming to improve overall accuracy for drug–ADR associations or severe outcomes [9, 10]. This paradigm overlooks a fundamental clinical reality: the same drug or regimen may exhibit markedly different risk drivers, risk magnitudes, and even risk rankings across sex subgroups. At present, there remains a lack of systematic modeling frameworks specifically designed to identify, interpret, and compare subgroup-specific risk landscapes.

This gap is particularly striking given that sex (along with age and other demographics) is widely recognized as a risk modifier [5, 6, 7, 13, 14]. Yet, many FAERS-based "intelligent" prediction pipelines do not treat sex heterogeneity as a central modeling target. Methodologically, "black-box" models such as deep learning and graph neural networks often provide limited transparency for explaining why risk differs between subgroups [15]. Even when sex is included as an input feature, its contribution may be masked or confounded by stronger predictors [7], consistent with observations reported by [10]. As a result, these models frequently cannot answer a key clinical question: *How the risk of a given adverse reaction differs between males and females for a specific drug?* This limitation hinders the delivery of actionable, trustworthy, and directly interpretable sex-differentiated safety insights for personalized risk prediction.

Finally, there is a notable lack of comparative validation using different methodological approaches to address the aforementioned question. To our knowledge, few studies have directly compared interpretable modeling paradigms—such as parametric linear models versus nonparametric ensemble learners—in their ability to characterize sex heterogeneity in ADR risk. Such comparisons are essential for evaluating the stability, depth of interpretability, and clinical usability of pharmacovigilance systems.

To address the aforementioned gaps, this study leverages FAERS data to develop an interpretable, sex-focused pharmacovigilance framework for adverse drug reactions by applying and comparing two machine learning approaches that are both predictive and interpretable at the subgroup level: (i) Relaxed LASSO regression with explicit sex–drug interaction terms, and (ii) sex-stratified Random Forest modeling combined with SHAP (SHapley Additive exPlanations) explanations. Within the generalized linear model framework, the Relaxed LASSO approach explicitly incorporates sex–drug interaction terms and employs a two-stage relaxation fitting strategy to refine coefficient estimation, thereby providing direct and quantitative estimates of sex-specific effects (e.g.,

coefficients and odds ratios) that are consistent with traditional clinical interpretation. Conversely, the Random Forest approach adopts a sex-stratified modeling strategy, leveraging bagging mechanisms to capture complex interaction patterns, and subsequently applies SHAP values to uniformly quantify feature contributions, offering interpretable insights at both the global and individual-case levels.

The core contribution of this study lies not only in constructing models under two distinct paradigms to reveal sex-specific drug risk profiles, but also in conducting cross-validation and systematic comparison of their results. We focus on the following: (1) the consistency of high-risk drug lists for males and females identified by different methods; (2) whether the conclusions regarding effect direction and magnitude from linear and nonlinear models are mutually complementary–If linear and nonlinear models learn different useful information, they may provide complementary evidence for risk assessment; and (3) the trade-offs among model interpretability, stability, and clinical operability. Through this multi-method comparative approach, we aim to elucidate the applicability scenarios and relative value of different interpretable machine learning paradigms in subgroup heterogeneity analysis using FAERS data. Overall, this work offers a preliminary methodological basis and empirical insights that could inform future efforts to develop more transparent and precise sex-stratified drug safety early warning systems.

2 Data Description

In the field of drug safety monitoring, pharmacovigilance research based on spontaneous reporting databases has become a core method for detecting adverse drug reactions (ADRs). As a publicly available database, the FDA Adverse Event Reporting System (FAERS) [8] provides an ideal platform for analyzing the complex association patterns between drugs and adverse reactions.

FAERS acts as a cornerstone for post-marketing drug safety surveillance. The database has been publicly available with quarterly updates since 2004, and as of 2024, it began providing daily updates via a public dashboard. It collects reports on adverse events, medication errors, and product quality complaints regarding marketed prescription drugs and therapeutic biologics. Data collection relies primarily on two methods: direct and indirect reporting. Direct reporting allows healthcare professionals and consumers to submit reports voluntarily via the FDA's MedWatch program (online, by mail, or by fax). Indirect reporting refers to adverse event reports submitted by pharmaceutical manufacturers based on their legal obligations after receiving information from healthcare professionals or consumers. In practice, approximately 75% of reports are submitted through manufacturers, resulting in a known degree of missing values in data fields.

This study utilizes public data from FAERS as its foundation. We batch-retrieved ADR reports from 2015 to 2024 via the API provided by the OpenFDA platform (`https://open.fda.gov/apis/drug/event/`). The dataset includes patient demographics (sex, age, weight), therapeutic information (drug names, active ingredients, and records of polypharmacy), and adverse reaction

information (standardized terminology and records of multiple simultaneous reactions). The data is stored in JSON format and organized by quarter. Each record represents an individual case safety report, with the total dataset covering over 13.71 million original records.

As our objective is to predict adverse drug reactions, it is necessary to extract drug information and corresponding adverse reactions from this raw data, followed by data preprocessing and further sample selection.

3 Data Preprocessing

To ensure data quality and align with research requirements, a systematic data preprocessing pipeline was implemented. Its overall logic follows a progressive sequence from an overview of the raw data to data cleaning and filtering, and finally to feature and sample selection.

3.1 Data Overview

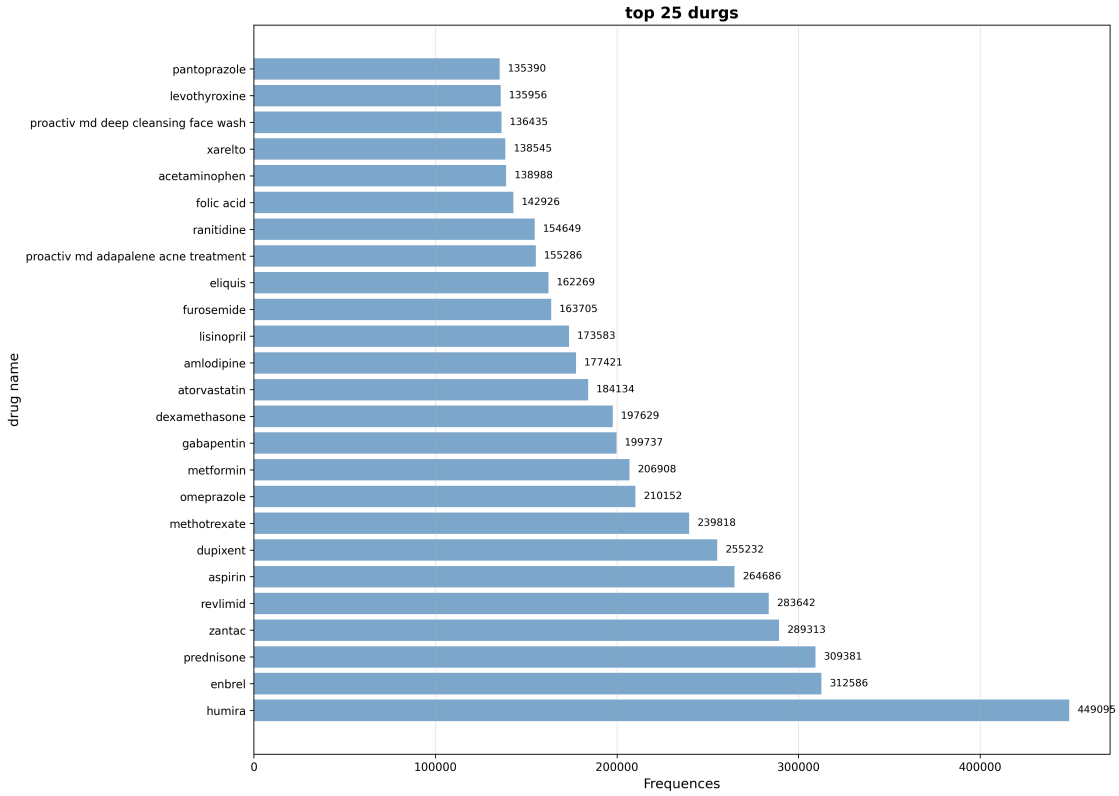

Figure 1: Frequencies of the 25 most frequently reported drugs

Using the FAERS public data obtained via the aforementioned API, we iteratively read all quarterly JSON data files from 2015-2024, extracting key fields and constructing structured records. Before applying cleaning or filtering to the initial dataset of over 13.71 million records, we performed an exploratory analysis of the raw data distribution.

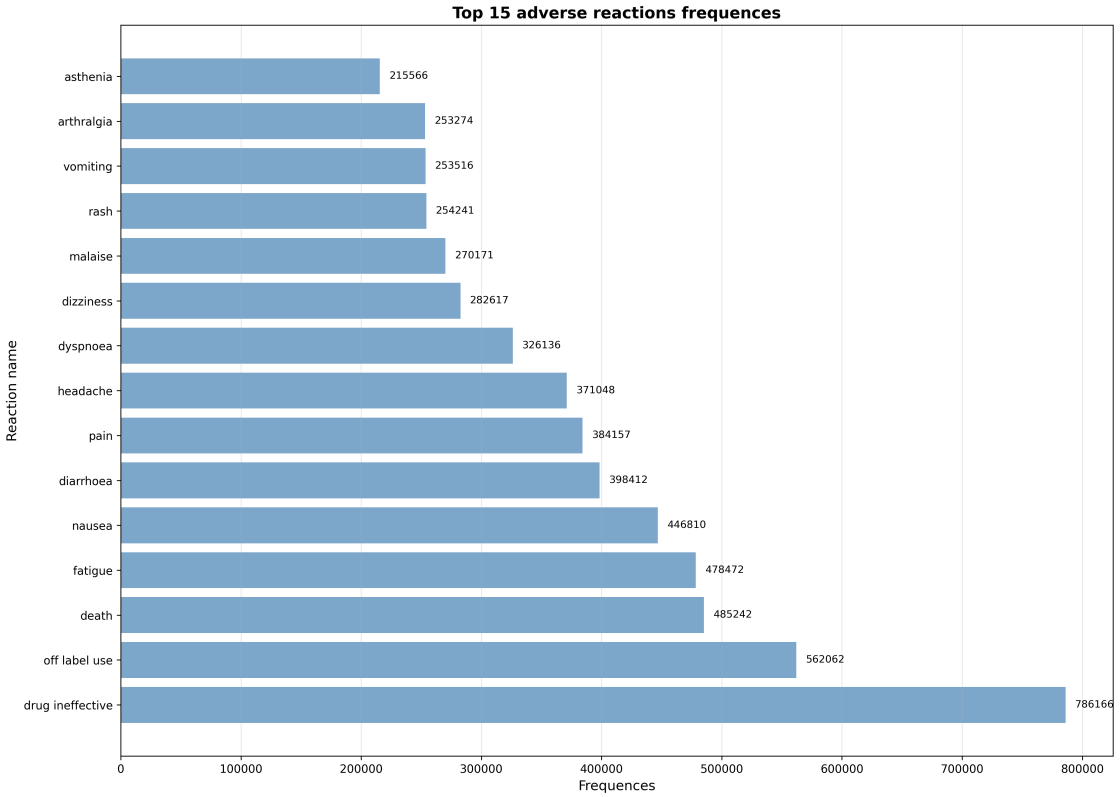

Figure 2: Frequency of reported adverse reactions

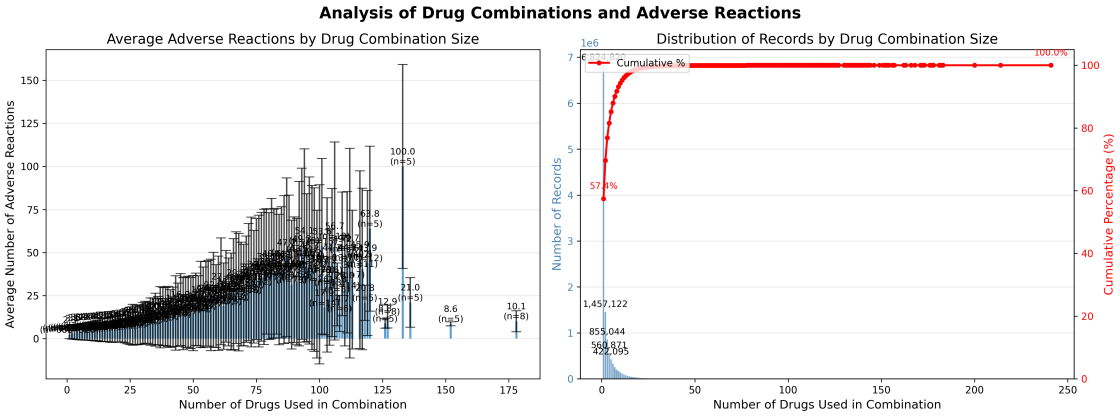

Figure 3: Relationship between the number of drugs in a combination and the number of reported adverse reactions

Note: Drug combination size refers to the number of drugs included in a single ADR report.

From the drug perspective, the dataset covers major therapeutic categories such as antibiotics, anticoagulants, cardiovascular drugs, antidepressants, antiepileptics, and chemotherapeutic agents. According to the statistics, the top 5% most frequently reported drugs appeared in 99.33% of all records. Figure 1 shows the 25 most frequently occurring drugs in the data, accounting for 14.4% of the total number of drug occurrences.

From the perspective of adverse reactions, the statistical results shown in Figure 2 indicate that the most frequently reported adverse reaction is "drug ineffective" with a frequency of 786,166 reports, significantly higher than other categories. This is followed by "off label use" and "death", with frequencies of 562,062 and 485,242, respectively, reflecting the severity of clinical medication risks. Regarding systemic and neurological symptoms, "fatigue", "nausea", "diarrhoea", "pain", "headache", and "dyspnoea" are all among the high-frequency categories, with frequencies ranging from approximately 320,000 to 480,000 reports. In addition, adverse reactions such as dizziness, malaise, rash, vomiting, arthralgia, and asthenia also appear in the high-frequency sequence, with relatively uniform distributions concentrated between 210,000 and 280,000 reports. Overall, the high-frequency adverse reactions span multiple systems including digestive, neurological, and musculoskeletal symptoms. Moreover, the elevated reporting rates of "drug ineffective" and "death" underscore the necessity of prioritizing drug quality control and the prevention of life-threatening risks in subsequent drug safety evaluations.

Additionally, we analyzed the relationship between drug combination complexity and the number of adverse reactions. As shown in Figure 3, as the number of drugs involved in a report increases, the average number of reported adverse reactions also generally shows an increasing trend. This preliminarily confirms the clinical observation that polypharmacy may increase the risk of adverse reactions. Moreover, the vast majority of records involve a relatively small number of concomitant medications: cases with fewer than five drugs account for over 80% of all records.

3.2 Data Cleaning and Filtering

After understanding the overall picture of the data, we implemented a data cleaning and filtering process to ensure data quality and the reliability of subsequent analyses. First, record uniqueness was verified based on the report ID. Subsequently, drug information was sorted. We then systematically assessed the completeness of each feature field.

As shown in Table 1, statistics revealed significant differences in the missing rates across different feature fields, which directly influenced the selection of available features for subsequent modeling. Due to excessively high missing rates, the age and weight fields were excluded from the predictive models, as retaining them would have required either massive sample deletion or the introduction of significant imputation bias. In contrast, the drug and adverse reaction fields were largely complete and were the core objects of analysis. The sex field had a relatively acceptable missing rate; therefore, it was retained as a key demographic feature for subgroup analysis. Figure 4 and Figure 5 show the sex missing rates of different drugs and different reactions, respectively. Both rates

fluctuate randomly within a small range of around 13.35%. Following previous works [10, 16], we further removed invalid records containing missing values in selected key fields (sex, drug information, adverse reactions). This progressive cleaning process resulted in a final analytical dataset of 11,886,107 clean records, filtered from the initial 13,717,173 raw reports.

Table 1: Missingness in the original feature fields

Field	Missing Count / Total Count	Missingness
sex	1,831,066 / 13,717,173	13.35%
age	6,027,051 / 13,717,173	43.94%
weight	11,279,506 / 13,717,173	82.23%
drugs	0 / 13,717,173	0.00%
reactions	4 / 13,717,173	≈0.00%

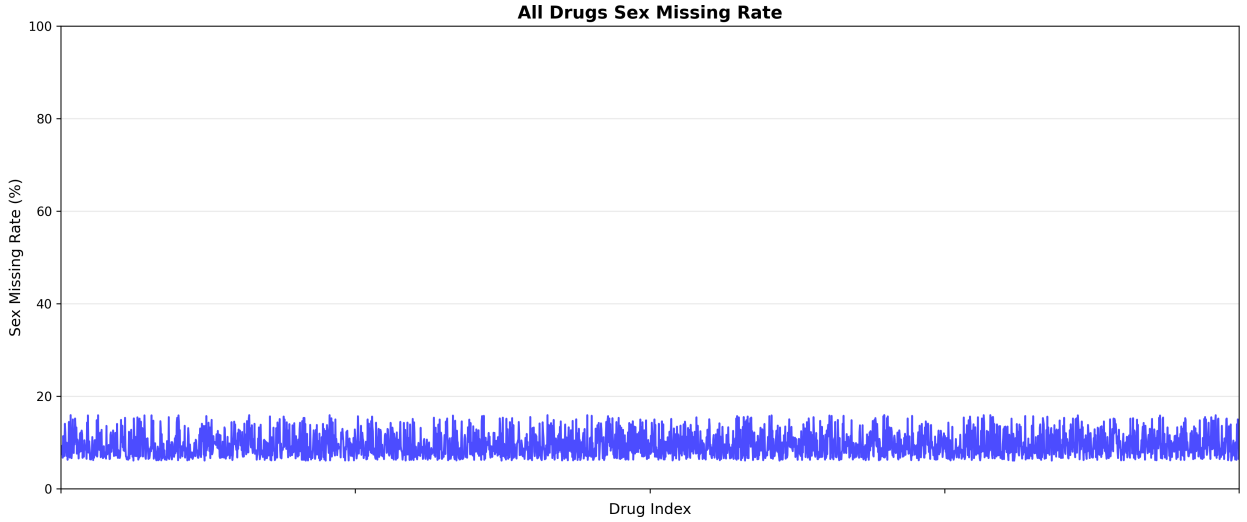

Figure 4: Missing rate of sex across drugs

3.3 Feature and Sample Selection

To systematically identify drugs significantly associated with the target adverse reactions from the cleaned data and to construct features for the machine learning models, this study employed a Disproportionality Analysis (DA) method for further feature selection.

As a data mining technique, it quantitatively detect potential "drug-adverse reaction" signals within spontaneous reporting system data. It assesses the association strength of drug-adverse reaction combinations by determining whether there is a statistical "disproportionality" between the reporting proportion of a specific adverse reaction for a drug and the background proportion, thereby indicating potential risks.

For Disproportionality Analysis, we utilized commonly used Proportional Reporting Ratio (PRR) method [17] for subsequent model building. For each target adverse reaction, the analysis proceeds

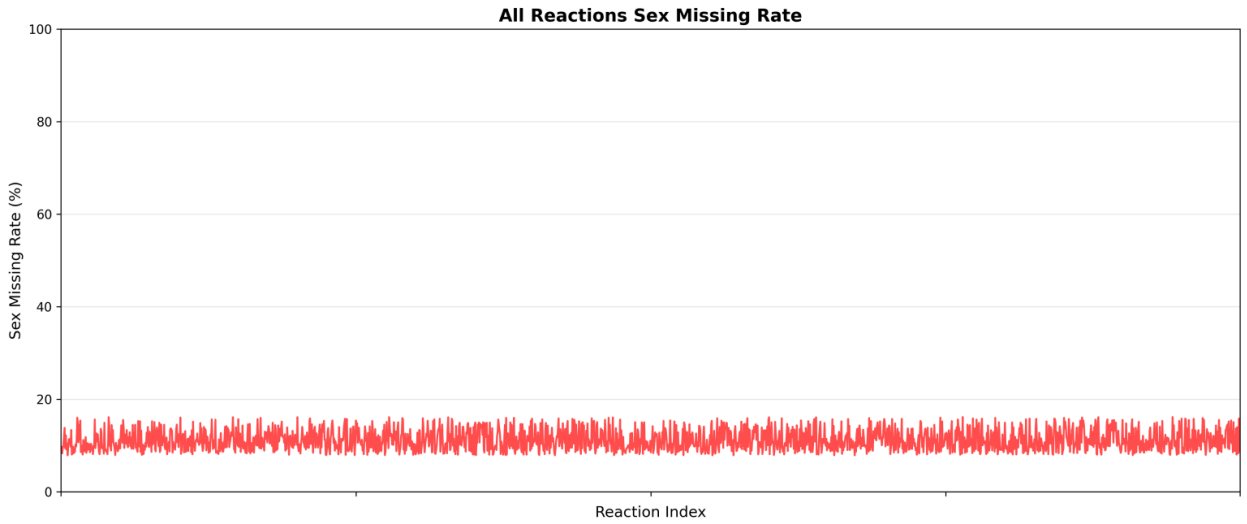

Figure 5: Sex Missing Rate of different reactions

as follows: First, the exposed population was identified by screening all reports containing that adverse reaction A. Then, all drugs $(x, y, ...)$ present in the exposed population were enumerated, and the frequency of drug and adverse reaction co-occurrence was counted. For each drug-adverse reaction combination, a 2x2 contingency table (Table 2) was independently constructed for comparison, considering the four dimensions of the target drug, other drugs, the target reaction, and other reactions. The corresponding PRR result is calculated accordingly:

$$\mathrm{PRR}(A, x) = \frac{a/(a+b)}{c/(c+d)}$$

For the calculation results of each drug-adverse reaction combination, threshold values were preset according to the Evans Criteria [18] as $PRR \geq 2$, statistically significant result ($\chi^2 \geq 4$), and meeting the minimum report count criterion ($a \geq 3$). This yields potential drug safety signals, detecting whether the reporting of a specific "drug-adverse reaction" combination is disproportionately high within the entire database context.

Table 2: 2x2 contingency table for drug x and adverse reaction A

	Target Reaction A	Other Reactions	Total
Target Drug x	a	b	a+b
Other Drugs	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

We specifically focused on an in-depth analysis of the 13 most frequent adverse reaction types, with results summarized in Table 3. Figure 6 visualizes the Disproportionality Analysis results for the 9 most frequent adverse reactions, highlighting significant variations in association strength

between different adverse reactions and drugs. Certain adverse reactions (e.g., "death", "pain") exhibited particularly strong associations with specific drugs. These findings provide a solid statistical and feature engineering foundation for building subsequent machine learning models.

Table 3: Disproportionality analysis results for drugs associated with the 13 most common adverse reactions

Adverse Reaction	Number of Risk Drugs	Total Records of Risk Drugs	Records Containing Both Risk Drug and Reaction	Coverage
death	768	1,925,185	287,588	14.94%
fatigue	2,058	3,447,431	274,482	7.96%
nausea	2,178	2,972,164	254,537	8.56%
diarrhoea	2,207	3,120,917	245,439	7.86%
pain	1,586	3,709,921	258,415	6.97%
headache	1,835	3,027,075	196,416	6.49%
dyspnoea	2,449	3,254,567	206,126	6.33%
dizziness	2,049	2,743,486	141,846	5.17%
malaise	1,597	2,534,136	109,111	4.31%
rash	1,461	3,327,870	144,489	4.34%
vomiting	2,627	3,311,884	142,886	4.31%
arthralgia	1,633	3,899,846	174,016	4.46%
asthenia	2,599	3,464,289	122,940	3.55%

Note: The two non-clinical reactions, "drug ineffective" and "off-label use," were excluded from the top 15 adverse reactions. Coverage is defined as the number of records containing both the risk drug and the adverse reaction divided by the total number of records involving the risk drugs.

Finally, for each adverse reaction, all records containing at least one drug meeting the Evans Criteria were selected as the study cohort. Within these selected samples, reports containing the corresponding adverse reaction A were labeled as positive samples, while those without adverse reaction A were labeled as negative samples. By extracting these adverse reaction-related risk drugs and corresponding data records, a systematic drug-adverse reaction association knowledge base was established, providing an important feature foundation for subsequent predictive model construction.

4 Methods

4.1 Relaxed LASSO with Interaction Terms

To investigate complex and potentially nonlinear association patterns between drugs and adverse reactions—particularly how demographic variables like sex may interact with drug use to alter adverse reaction risk—this study employed the Relaxed LASSO (Least Absolute Shrinkage and Selection Operator) regression method with interaction terms [19, 20].

The Relaxed LASSO extends the standard LASSO in two stages to address the latter's inherent bias-variance trade-off. In the first stage, a standard LASSO with penalty parameter λ_1 performs variable selection, identifying a sparse set of predictors (including main effects and sex-drug interaction terms) associated with adverse reactions. In the second stage, a reduced penalty parameter λ_2

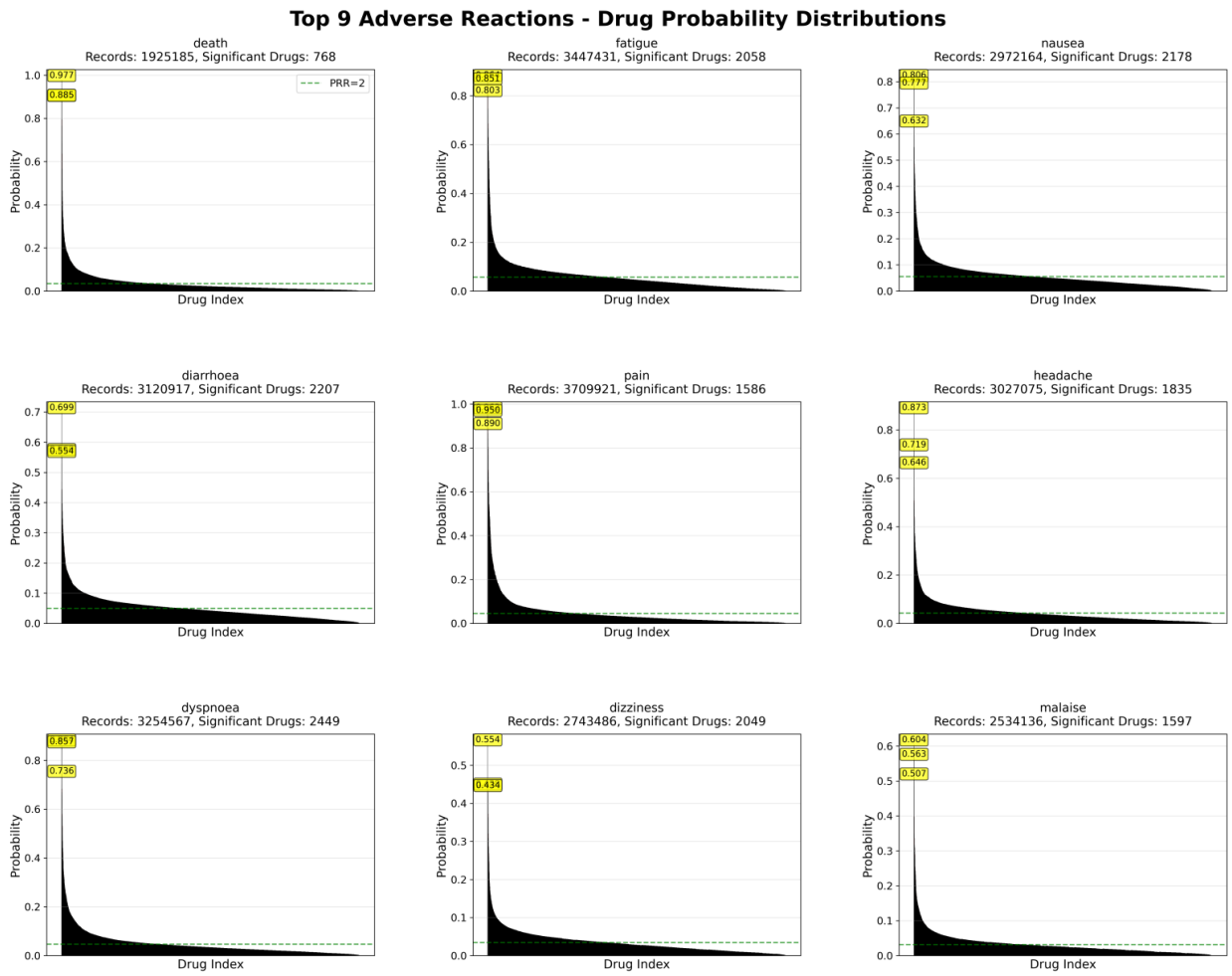

Figure 6: Disproportionality analysis results for high-frequency adverse reactions

(where $\lambda_2 \leq \lambda_1$) is applied exclusively to the selected variables, re-estimating their coefficients with less regularization. Cross-validation is utilized to determine the optimal λ_2 by minimizing prediction error (e.g., deviance) given the features retained in the first stage. By decoupling variable selection (controlled by λ_1) from coefficient shrinkage (controlled by λ_2), this approach retains LASSO's capacity for sparse selection in high-dimensional settings while yielding coefficient estimates with lower bias and improved stability. This provides more reliable quantification of sex-specific drug effects.

Traditional methods typically evaluate individual drug effects or perform isolated subgroup analyses, which cannot systematically quantify sex–drug interactions—i.e., whether the same drug presents significantly different risks for males and females. This study addresses this gap by constructing sex-drug interaction terms and applying Relaxed LASSO to achieve two objectives: first, to robustly select key features (main effects and interactions) associated with target adverse reactions in high-dimensional, sparse spontaneous reporting data; and second, to obtain more stable coefficient estimates with reduced bias by relaxing the penalty strength. This enables the precise

quantification of sex-based differences in drug effects, supporting nuanced drug safety surveillance.

This method operates within a logistic regression framework and identifies sex-specific signals through interaction terms and a two-stage LASSO procedure. Its core workflow proceeds as follows:

First, in addition to the feature selection from the data pre-processing stage, further feature construction and data preparation are performed. The set of significant drugs identified in the earlier Disproportionality Analysis serves as the base features. For each adverse reaction report, in addition to encoding the drugs ("drug_XXX" feature coded as 1 or 0) and sex ("sex_male": 1 = male, 0 = female), interaction terms between all significant drugs and the sex variable are constructed as "sex_male_x_drug_XXX". The coefficient of this interaction term directly represents the difference in drug effect between males and females. Given the highly imbalanced nature of the data, balanced subsampling was applied to the highly imbalanced data to ensure stability during model training. Specifically, we performed random downsampling of the majority class (i.e., non-events) to achieve a target event-to-non-event ratio of 1:1.

Second, the dataset is divided using a training–validation–test set split strategy, with the two-stage Relaxed LASSO modeling performed on the training. In the strict feature selection stage, a standard LASSO logistic regression model with a relatively strong penalty coefficient is fitted on the training set. The goal of this stage is to conduct preliminary aggressive feature screening, selecting a sparse subset S of features with non-zero coefficients from all drug main effects and sex–drug interaction terms. Then, in the relaxation refitting stage, using only the feature subset S selected in the first stage, a new LASSO model is refitted, with its reduced penalty parameter selected by cross-validation. This relaxation process aims to mitigate the coefficient estimation bias potentially introduced by the strong penalty in the first stage, thereby obtaining more reliable coefficient estimates on the selected feature subset. Model performance evaluations were conducted separately for males and females.

The final model performance is evaluated on independent validation and test sets using metrics including accuracy, ROC-AUC, PR-AUC, and F1 score to confirm model effectiveness. Accuracy measures the overall proportion of correct predictions, offering a general sense of model correctness. ROC-AUC assesses the model's discriminative ability across all classification thresholds, with higher values indicating better separation between positive and negative cases. PR-AUC is particularly valuable for imbalanced data, as it evaluates the trade-off between precision and recall. The F1 score, the harmonic mean of precision and recall, provides a balanced summary of model performance on the positive class, making it especially informative when class distributions are uneven.

Finally, further interpretation was conducted on the sex-specific effects obtained from the modeling process. For the non-zero coefficients in the final model, the coefficient of "drug_XXX" represents the effect of the drug in the female baseline, while the coefficient of "sex_male_x_drug_XXX" represents the additional increase or decrease in the drug effect in males compared to females. By adding the main effect coefficient of the same drug to its interaction term coefficient, the total effect of the drug in the male subgroup is obtained. sex-specific drug risk rankings and complete

effect sizes were calculated separately for both males and females, yielding key sex–drug interaction effects.

Through the above Relaxed LASSO modeling process with interaction terms, this study effectively combines interaction analysis with high-dimensional feature selection, robustly identifies important signals while avoiding overfitting, and transforms the hypothesis testing of sex differences into statistical inference on the coefficients of interaction terms, producing clear and intuitive results. The final sex-specific drug weight results can be directly used in subsequent sex-differentiated risk prediction modeling or to provide clinicians with prioritized monitoring drug lists tailored to different sexes. This method offers a powerful data analysis tool for discovering and validating sex heterogeneity in drug safety.

4.2 Random Forest

To independently investigate drug–adverse reaction risk patterns in male and female populations separately, and to overcome the limitations of linear models in capturing complex real-world effects such as multi-drug co-occurrence and synergistic biological pathways, this study constructed separate Random Forest (RF) models for males and females and applied the SHAP (SHapley Additive exPlanations) framework for model interpretation and feature importance quantification.

Although a single model that includes sex as a covariate can provide a general analysis, it struggles to reveal potentially entirely different risk contribution structures and interaction networks for the same drug across sex groups. This approach pursues two core objectives: first, to leverage the ensemble decision-making capability of Random Forest [21] to build high-performance ADR risk identification models separately for males and females, thereby capturing potential sex heterogeneity; second, to use the cooperative-game-theory-based SHAP framework [22, 23] to quantify and visualize the contribution magnitude and direction of each drug feature to the predictions in the male and female models respectively. This effectively breaks the "black box" nature of the model and transforms predictions into interpretable, verifiable sex-specific lists of risk-associated drugs with ranked importance, thus providing data-driven insights for sex-differentiated pharmacovigilance.

The method follows a "model construction – performance evaluation – interpretability mining" workflow that systematically integrates prediction and explanation. Utilizing the same dataset described in the Relaxed LASSO section, the data was stratified by sex into two independent subsets. For each subset, a feature matrix was constructed based on whether the patient was exposed to each significant drug. Given the highly imbalanced nature of the data in each sex-specific subset, balanced subsampling was applied following the same approach used for the Relaxed LASSO described in the previous section. Also, each sex-specific dataset was then randomly split in a stratified manner into 70% training, 15% validation, and 15% test sets. For comparison purposes, we also conducted supplementary experiments without balanced subsampling, with results provided in the Appendix.

Subsequently, independent RF models were developed and validated on the male and female data subsets separately. These models adopt an ensemble learning strategy: multiple diverse de-

cision trees are constructed via bootstrap sampling. At each node split, a random subset of drug features is evaluated to determine the optimal split point, thereby enhancing model robustness. Hyperparameters, such as the number of trees and maximum tree depth, were optimized to maximize predictive performance. RF models were evaluated using the same methodology applied to Relaxed LASSO, as detailed in the previous section.

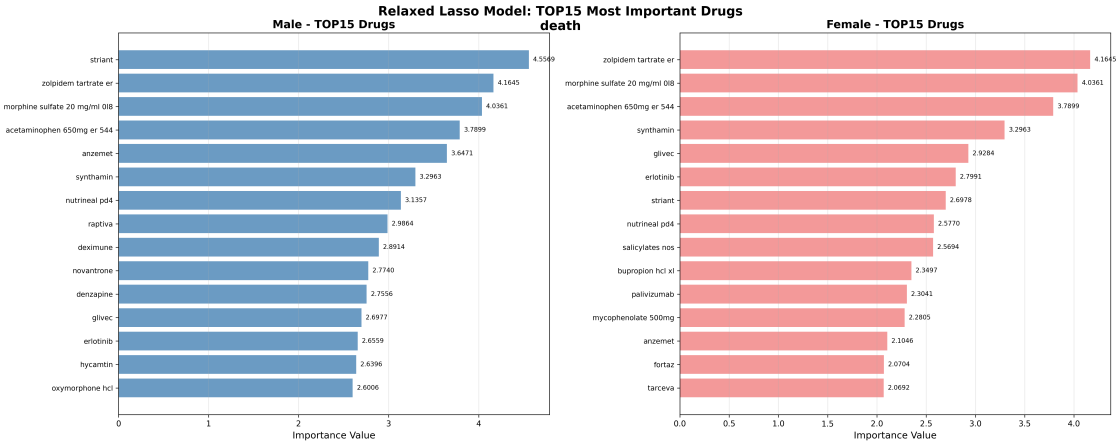

Figure 7: Sex-specific drug effect estimates from the Relaxed LASSO model for "death"

Finally, sex-specific drug importance was interpreted using the SHAP framework applied separately to each trained sex-specific model. SHAP values were computed for every sample in the training set, and the mean of their absolute values was calculated to obtain the global importance score for each drug feature within the corresponding sex model. Unlike traditional Gini-based importance, SHAP values are consistent and can characterize the direction of influence. Furthermore, the raw SHAP importance scores were normalized to facilitate a direct comparison of rankings between the male and female models. The system outputs the normalized Top 15 key drug rankings and generates distinct risk drug lists for each sex. By identifying both shared high-risk drugs and sex-specific high-risk drugs, this provides an intuitive visualization of sex differences in drug risk patterns.

5 Results

This study employed two distinct analytical frameworks—Relaxed LASSO regression with interaction terms and Random Forest combined with the SHAP framework—to model and analyze three prevalent adverse drug reactions ("death", "fatigue", and "nausea") using FAERS data. The primary objective was to systematically compare the predictive performance of these two methods across sex subgroups, and to examine the similarities and differences in the sex-specific key risk drug patterns identified by each approach. The main results are presented below from three perspectives: (1) a comparative evaluation of model performance across sex subgroups; (2) the identification of sex-specific high-risk drugs; and (3) an analysis of the consistency of sex-preference signals between

the two modeling approaches.

Table 4: Sex-specific sample sizes for each adverse reaction

Adverse Reaction Type	Number of Drug Types	Gender	Number of Target ADR Samples	Total Sample
death	768	Male	159,481	934,337
		Female	128,107	990,848
fatigue	2058	Male	89,603	1,256,881
		Female	184,879	2,190,550
nausea	2178	Male	63,136	1,013,936
		Female	191,401	1,958,228

5.1 Sex-Subgroup Model Performance

The predictive performance of the two modeling approaches for adverse drug reactions across sex subgroups is summarized in Table 5. Overall, across all three adverse reaction prediction tasks, the Relaxed LASSO model consistently outperformed the Random Forest + SHAP approach in terms of accuracy (ACC), ROC-AUC, PR-AUC, and F1 score. This demonstrates its comprehensive advantages in capturing sex–drug interaction effects and performing high-dimensional feature selection. After averaging model performance across male and female subgroups, the Relaxed LASSO model achieved mean improvements of 0.0280 in ACC, 0.0240 in ROC-AUC, 0.0270 in PR-AUC, and 0.0652 in F1 score over the Random Forest + SHAP approach across the three tasks. While the Random Forest + SHAP model offered distinct strengths in interpretability, its predictive efficacy was comparatively lower.

From the perspective of adverse reaction types, both models achieved higher predictive performance (AUC generally > 0.75) for more clinically severe outcomes—namely death—than for symptomatic reactions such as "fatigue" and "nausea" (AUC approximately 0.60–0.70). This discrepancy may be attributed to the more clearly defined medical decision-making processes or pathological mechanisms underlying the former, whereas the latter are subject to greater subjectivity in reporting and a wider range of contributing factors.

At the sex subgroup level, both models revealed performance disparities between males and females, although the specific patterns varied depending on the adverse reaction type and the modeling method employed. For instance, using the Relaxed LASSO model, the ROC-AUC for "fatigue" was notably higher in the male subgroup (0.7094) than in the female subgroup (0.6590). These findings suggest that predictive efficacy may be imbalanced across sex groups within the same model. Moreover, the subgroup exhibiting superior predictive performance was not consistently aligned across modeling approaches. For example, in predicting "death", the Relaxed LASSO model yielded a higher F1 score in males (0.7247) than in females (0.6663), whereas the Random Forest model achieved higher ROC-AUC (0.7532), PR-AUC (0.7563), and F1 (0.6353) in females. For "fatigue", the Relaxed LASSO model performed better in the male subgroup (with a ROC-AUC

Table 5: Predictive performance of Relaxed LASSO and RF models across sex subgroups

Adverse Reaction Type	Modeling Method	Male				Female			
		ACC	ROC-AUC	PR-AUC	F1	ACC	ROC-AUC	PR-AUC	F1
Death	Relaxed LASSO	0.6956	0.7719	0.7975	0.7247	0.7013	0.7737	0.7596	0.6663
	RF+SHAP	0.6655	0.7421	0.7536	0.5832	0.6750	0.7532	0.7563	0.6353
Fatigue	Relaxed LASSO	0.6634	0.7094	0.6942	0.6715	0.6113	0.6590	0.6640	0.5679
	RF+SHAP	0.6362	0.6713	0.6583	0.6114	0.5900	0.6430	0.6410	0.5053
Nausea	Relaxed LASSO	0.6369	0.6658	0.5972	0.5089	0.6072	0.6413	0.6531	0.6704
	RF+SHAP	0.5983	0.6329	0.6171	0.5393	0.5830	0.6351	0.6369	0.4441

0.0504 higher), while the Random Forest model showed only marginal sex differences. These results indicate that different modeling approaches capture distinct sex-specific risk patterns, and that a single modeling strategy may be insufficient to fully elucidate the multidimensional nature of sex heterogeneity in adverse drug reaction risk.

5.2 Sex-Specific Key Risk Drugs

The two methods revealed, from different analytical perspectives, sex-specific risk drug profiles for each adverse reaction. First, the Relaxed LASSO model was used to extract and compare the top 15 most important drug features in predicting "death" across sex subgroups. As shown in Figure 8, the experimental results indicate both divergence and convergence in the key predictive drug factors between males and females. Specifically, in the male model, the most important drug was Striant, with an importance score of 4.5569. Conversely, in the female model, the top-ranked drug was Zolpidem Tartrate ER, with an importance score of 4.1645, followed closely by Morphine Sulfate and Acetaminophen, suggesting that sleep-regulating and analgesic agents play a more central role in risk assessment for females. Despite these differences, considerable overlap was observed between the top 15 drugs in both sexes. Drugs such as Zolpidem Tartrate ER, Morphine Sulfate, and Glivec (a tyrosine kinase inhibitor) exhibited consistently high importance across both subgroups, indicating that certain sedatives, analgesics, and antineoplastic agents have universal risk irrespective of sex.

Expanding upon this, we focused on comparing the relative importance of the same drugs across sex subgroups within the model and visualized the results. In a normalized comparative plot of 50 sampled drugs (Figure 9), the model revealed a high degree of cross-sex correlation (Spearman's p = 0.8535, Pearson's r = 0.8632), indicating that linear risk factors remain relatively stable between males and females. Despite this overall consistency, certain drugs exhibited marked sex-specific preferences. For instance, Lupron depot (a GnRH agonist), Myfenax (an immunosuppressant), and Denzapine (an antipsychotic) showed substantially higher normalized importance (scaled to 0–1 via min-max normalization) in males than in females, with differences of +0.487, +0.339, and +0.297, respectively. Conversely, Tamsulosin cr (an alpha-blocker), Viramune (an antiretroviral), and Prochlorper (an antipsychotic and antiemetic) demonstrated significantly greater risk contributions in females, with difference values of –0.328, –0.300, and –0.283, respectively.

In the Random Forest + SHAP model, results are derived from feature importance, and the

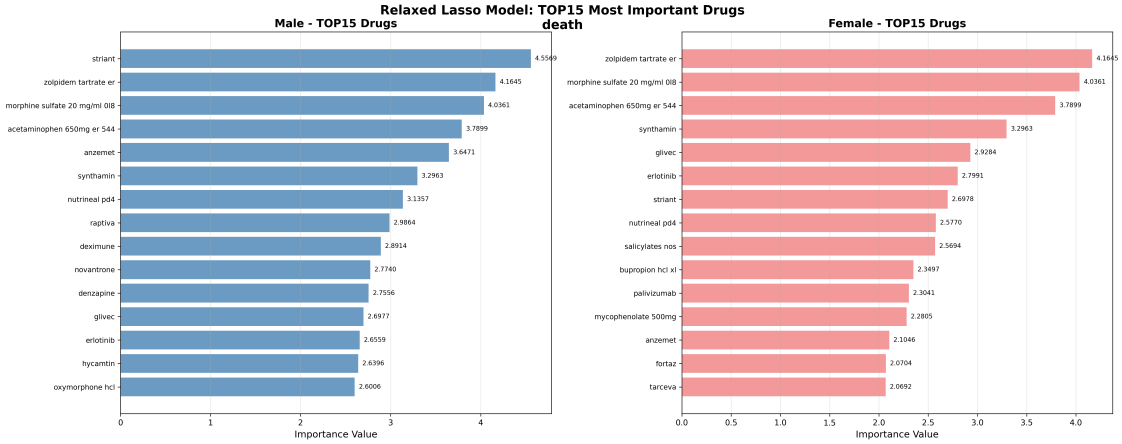

Figure 8: Sex-specific risk drug effects of the Relaxed LASSO model for the "death" adverse reaction

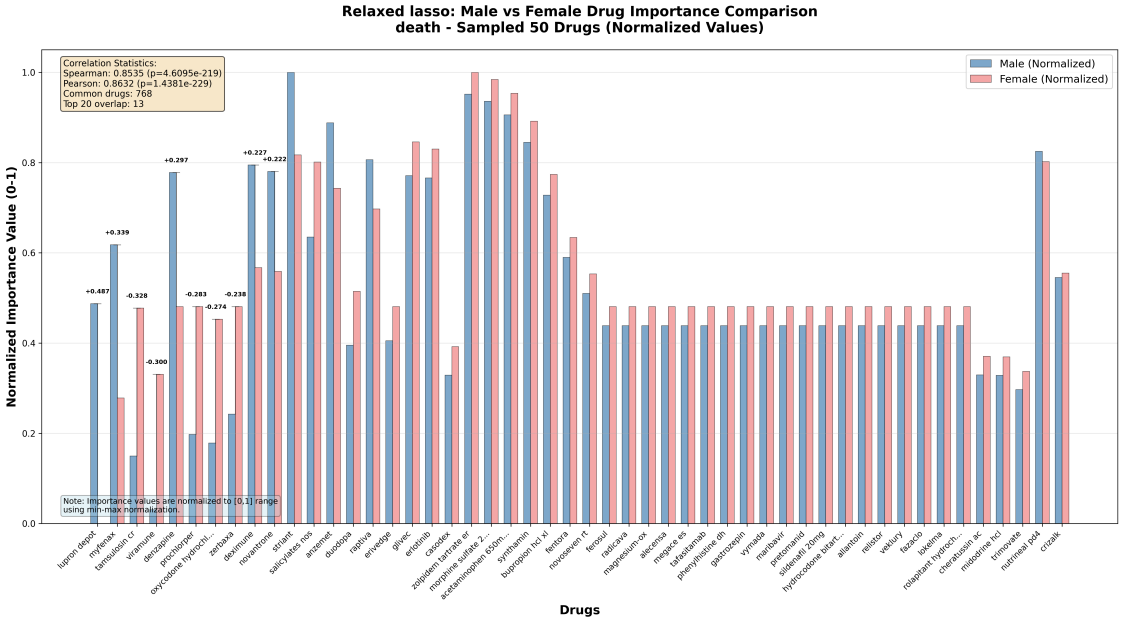

Figure 9: Comparison of sex-specific importance among 50 sampled drugs in the Relaxed LASSO model among "death"

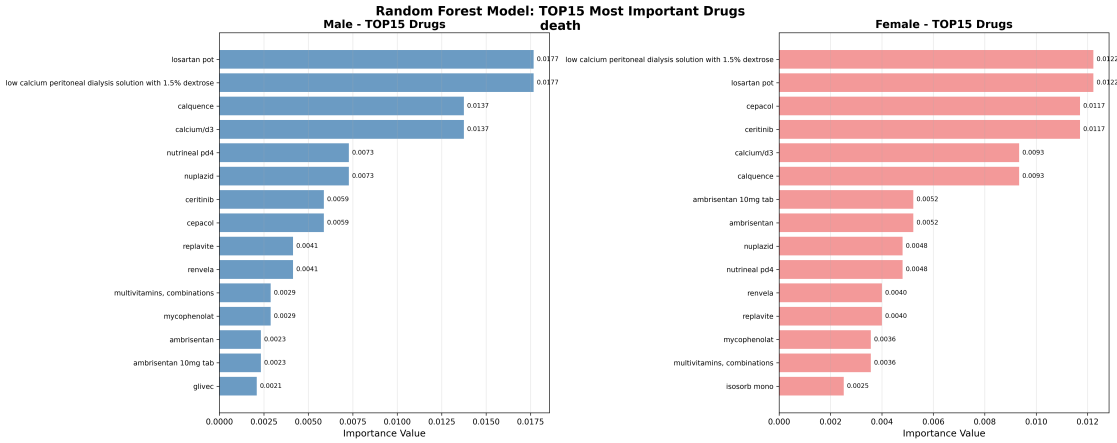

Figure 10: Sex-specific standardized SHAP values for risk drugs in the Random Forest model for "death"

top 15 drug lists reflect the contribution of each drug to the model's prediction of whether a target adverse reaction occurs within a given sex subgroup. The highest-ranked drugs represent key drivers of the model's predictive decisions. For example, in the case of "death" (Figure 10), the model identified losartan potassium and low calcium peritoneal dialysis solution with 1.5% dextrose as the most important drugs in both male and female models, with normalized SHAP contributions of 10.98% and 6.94%, respectively—both exceeding the 5% threshold. Furthermore, drugs such as Calquence, calcium/vitamin D3, Nutrineal PD4, and Nuplazid consistently appeared in the top five lists for both sexes, indicating that these agents are strong cross-sex predictors of mortality-related adverse reactions.

Overall, the Random Forest + SHAP approach revealed substantial overlap in the top 15 risk drugs identified across sex subgroups for each adverse reaction. For "death", for instance, 8 out of the top 10 drugs were shared between males and females, indicating a high degree of cross-sex consistency. However, while certain adverse reactions showed highly similar drug risk profiles across sexes, modeling results for other reactions exhibited pronounced sex-specific differences. These findings further underscore the necessity of conducting sex-stratified analyses in drug safety surveillance and risk factor investigation.

Additionally, further analysis of sex differences and correlation based on the Random Forest results (in Figure 11) reveals a very high Pearson correlation coefficient (0.9057) between male and female importance scores for the drugs, which is largely attributable to the strong consistency of top-ranked drugs such as losartan potassium and low calcium peritoneal dialysis solution. However, the Spearman correlation coefficient was relatively lower (0.6324), indicating notable ranking discrepancies among moderately and lower-ranked drugs. Specifically, the normalized importance of cepacol and ceritinib is substantially higher in females (≈ 0.958) than in males (≈ 0.332), both showing a difference value of –0.625. Likewise, ambrisentan exhibits greater feature importance in females, with a difference value of approximately –0.295. These disparities highlight the existence

of sex-specific driving strength for certain drugs within nonlinear risk contribution structures.

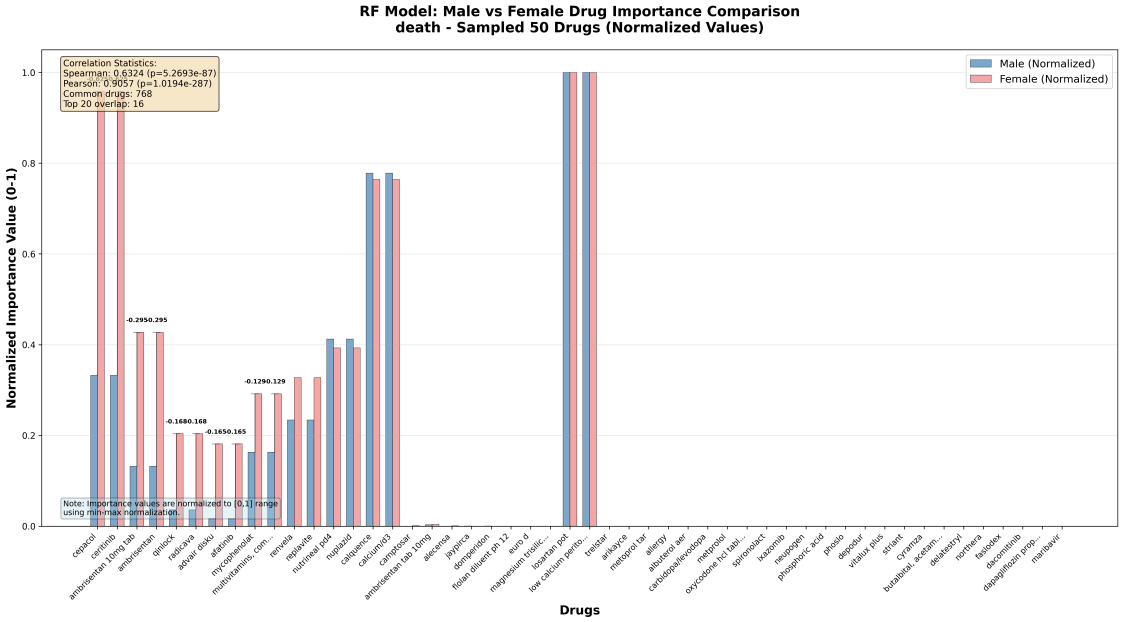

Figure 11: Comparison of sex-specific importance among 50 sampled drugs in the Random Forest model for "death"

The two methodological approaches described above depict distinct risk drug profiles for adverse reactions. Relaxed LASSO focuses on identifying specific drugs—particularly those with pronounced sex-based differences, such as hormonal agents—while Random Forest combined with SHAP emphasizes highly consistent cross-sex core predictors. This divergence highlights the substantial influence of methodological choice on the identification of "key risk drugs." Relaxed LASSO excels at detecting and quantifying extreme differential signals, whereas SHAP analysis based on Random Forest models more comprehensively reflects the feature systems underlying model decisions. The conclusions drawn from the two approaches are thus complementary in nature.

Similar methodological contrasts were observed in the analysis of other adverse reactions. For "fatigue", the results from the Relaxed LASSO model (Figure 12) show that the drug with the highest risk contribution in males is Vicodin HP (an opioid analgesic), with an importance score of 3.3617. In females, the drug with the highest importance is Lenvima (an anticancer agent), with a score of 3.1342. Lenvima also appears in the male top 15 drug list (3.1342). In addition, the male and female risk drug lists share several overlapping agents, including hydrocortisone acetate (3.0145), olaparib (2.9715), tildin (2.6986), and calcium 500+D (2.5649). Despite considerable overlap across sexes for drugs such as Lenvima, hydrocortisone acetate, and olaparib, differences remain in the ranking and effect magnitude of certain specific drugs, reflecting sex heterogeneity in the risk-driving factors associated with fatigue-related adverse reactions.

We further focused on comparing the importance of the same drugs across sex subgroups in the Relaxed LASSO model for the adverse reaction "fatigue". In the normalized comparative plot

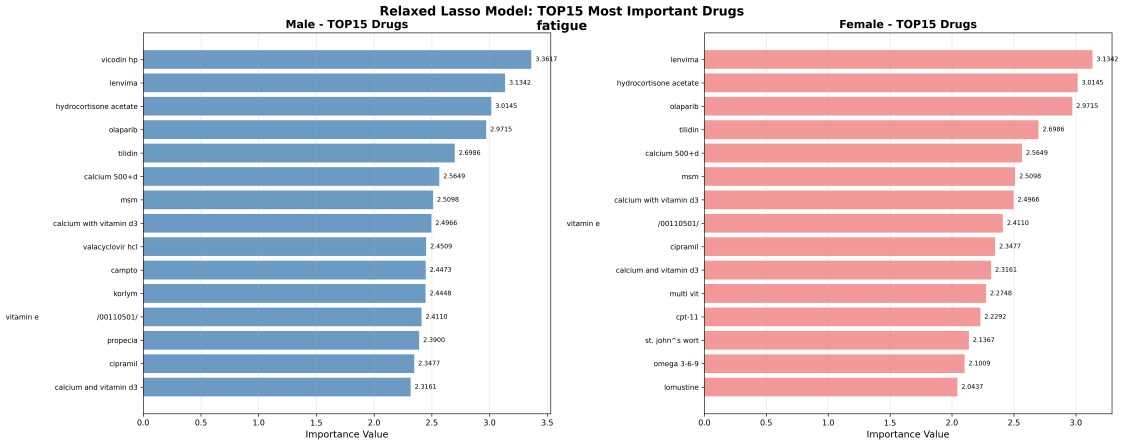

Figure 12: Sex-specific risk drug effect estimates from the Relaxed LASSO model for "fatigue"

of 50 sampled drugs (Figure 13), the model exhibited high correlation (Spearman's p = 0.8497, Pearson's r = 0.8754), indicating that linear risk factors are generally consistent in their trends between sexes. Nevertheless, several drugs displayed marked sex-specific preferences. For instance, the normalized importance of duloxetine hcl was substantially higher in females than in males (0.73 vs. 0.24, difference value = –0.488), whereas vicodin hp showed greater importance in males (1.0 vs. 0.56, difference value = +0.438). Senna s also had significantly higher weight in females (0.98 vs. 0.41, difference value = –0.433). Other drugs, such as zopiclone (0.99 vs. 0.40, difference value = +0.339) and dexamethason (0.97 vs. 0.38, difference value = –0.333), also exhibited notable sex differences.

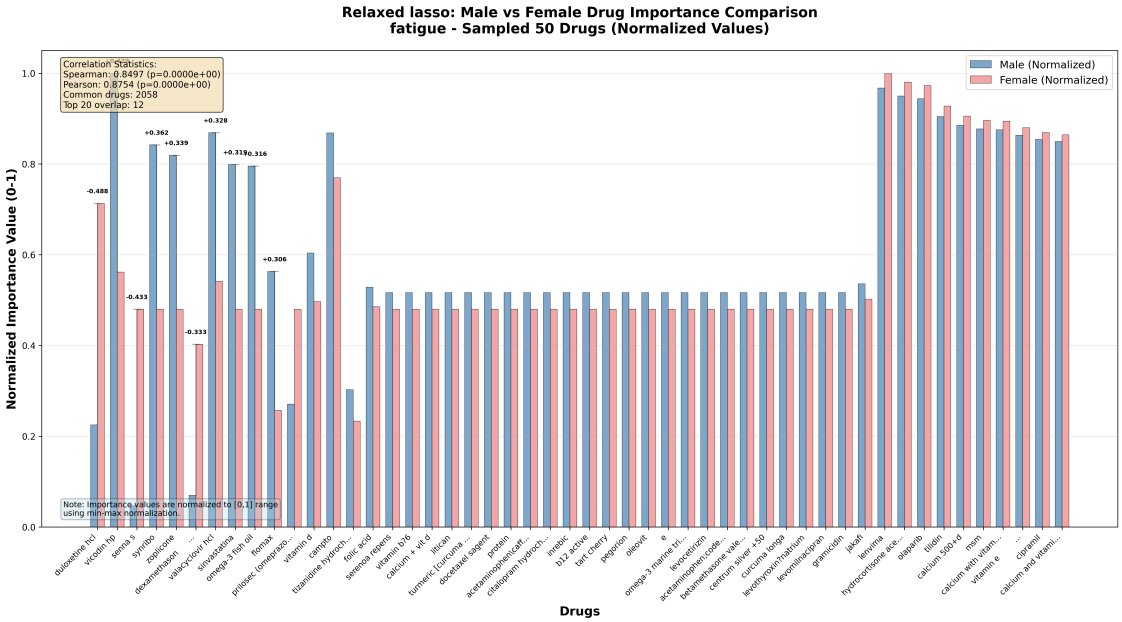

Figure 13: Comparison of sex-specific importance among 50 sampled drugs in the Relaxed LASSO model for "fatigue"

The analysis of the adverse reaction "fatigue" using the Random Forest + SHAP model (Figure 14) revealed a distinct risk profile. This model identified sulfadiazine/trimethoprim as the most important drug in the male model, with a normalized SHAP value of 0.0026, while marevan (warfarin) was the most important drug in the female model, with a normalized SHAP value of 0.0023. In addition, maprotiline, marevan (warfarin), and cholecalciferol (vitamin D3) consistently ranked among the top five drugs in both sexes, with SHAP values ranging from 0.0014 to 0.0023. Overall, the Random Forest + SHAP approach revealed a certain degree of overlap in the top 15 risk drugs for "fatigue" across sex subgroups, indicating that some risk factors are shared between males and females. However, pronounced sex differences were observed among the highest-risk drugs.

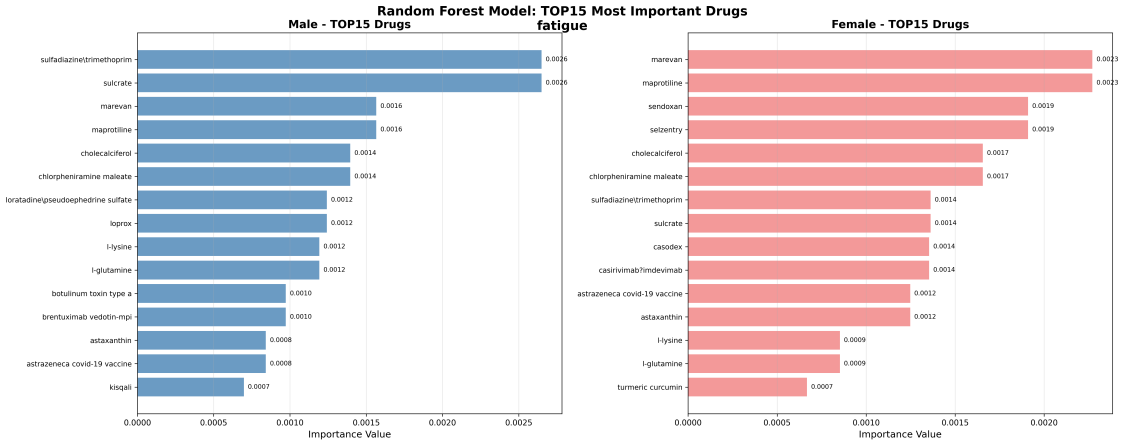

Figure 14: Sex-specific standardized SHAP values for risk drugs in the Random Forest model for "fatigue"

Additionally, in the further analysis of sex differences and correlation based on the Random Forest model results for "fatigue". Figure 15 shows a Pearson correlation coefficient of 0.7190 and a Spearman correlation coefficient of 0.5106, indicating a moderate overall correlation in drug importance between sexes, with more pronounced differences in ranking structure. The sampled data further revealed drugs with substantial sex-specific differences. For instance, the normalized importance of sendoxan (cyclophosphamide) and selzentry (maraviroc) was significantly higher in females than in males (difference value = –0.709). Similarly, casodex (bicalutamide) and casirivimab/imdevimab also exhibited a female-biased preference, with difference values of approximately –0.517. In contrast, loratadine/pseudoephedrine sulfate and loprox showed higher importance in males (difference value = +0.416).

For the adverse reaction "nausea", the results from the Relaxed LASSO model (Figure 16) show that the drug with the highest effect estimate in both males and females was magnesium glycerophosphate, with an effect value of 3.1606 in both sexes. The male top 15 drug list also included thiola (3.1425) and dulaglutide (3.1010), while the female list included dulaglutide (3.1010) and olaparib (2.9274). The top 15 drug lists for males and females exhibited substantial overlap and comparable effect estimates, indicating strong cross-sex consistency in the linear effects of most

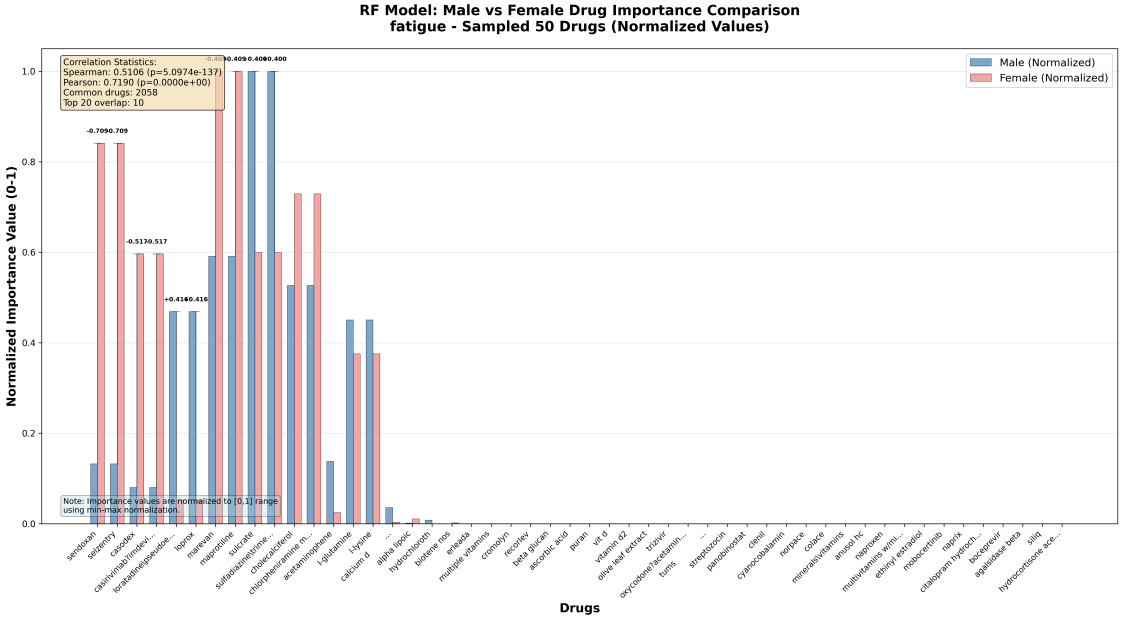

Figure 15: Comparison of sex-specific importance among 50 sampled drugs in the Random Forest model for "fatigue"

key risk drugs associated with nausea.

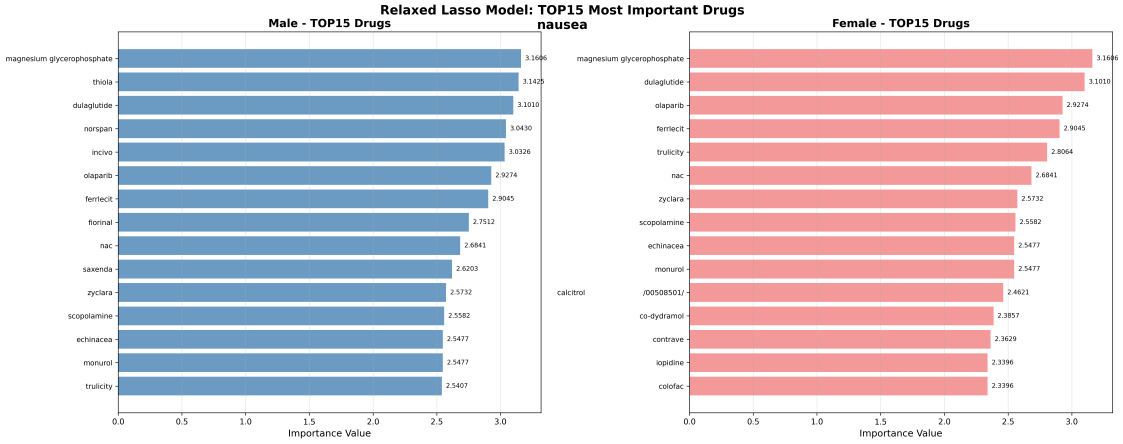

Figure 16: Sex-specific risk drug effect estimates from the Relaxed LASSO model for "nausea"

Expanding upon this, we compared the importance of the same drugs across sex subgroups in the Relaxed LASSO model for the adverse reaction "nausea". In the normalized comparative plot of 50 sampled drugs (Figure 17), the model exhibited high cross-sex correlation (Spearman's p = 0.8264, Pearson's r = 0.8329). Nevertheless, certain drugs still demonstrated notable sex differences. For example, exenatide showed substantially higher normalized importance in females than in males (difference value = –0.613), and arava (leflunomide) also had greater weight in females (0.88 vs. 0.48, difference value = –0.426). In contrast, norspan and incivo exhibited significant male-biased preferences, with difference values of +0.488 and +0.439, respectively.

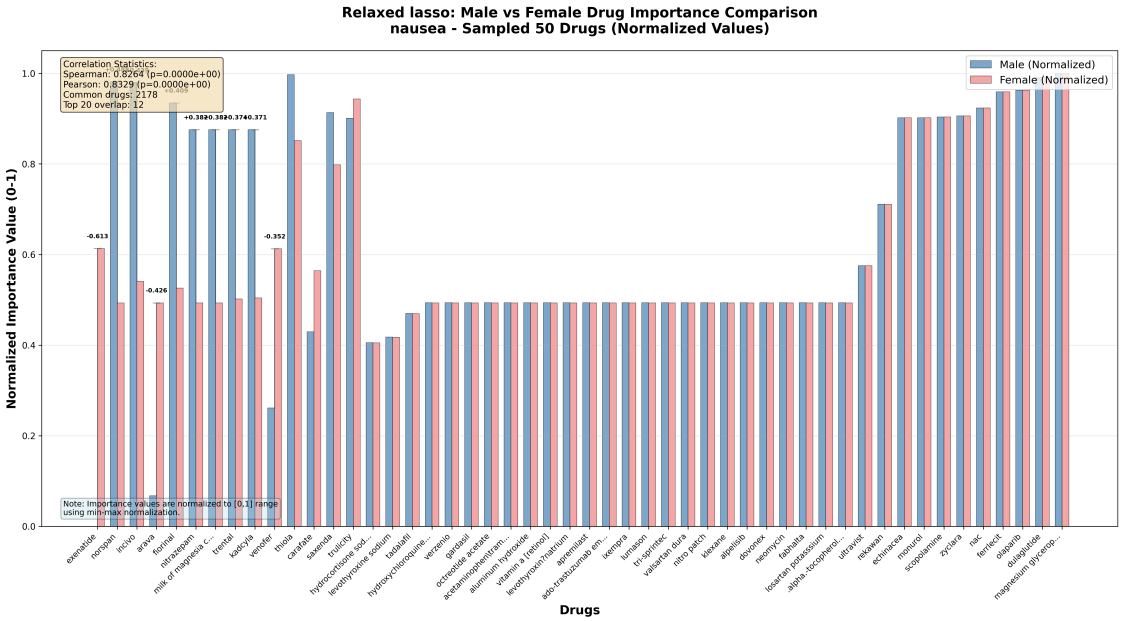

Figure 17: Comparison of sex-specific importance among 50 sampled drugs in the Relaxed LASSO model for "nausea"

The analysis of the adverse reaction "nausea" using the Random Forest + SHAP model (Figure 18) revealed a distinct risk profile. This model identified Influenza virus and Influenza vaccine as the most important drugs in the male model, both with a normalized SHAP value of 0.0013. In the female model, the most important drugs were omega-3 and omega-3-6-9, which also appeared in the male model, both with normalized SHAP values of 0.0009. Notably, there was almost no overlap in the top 15 risk drug lists between sexes for "nausea". Overall, the Random Forest + SHAP approach demonstrated that the core drug features driving predictions in the "nausea" model differ substantially between males and females.

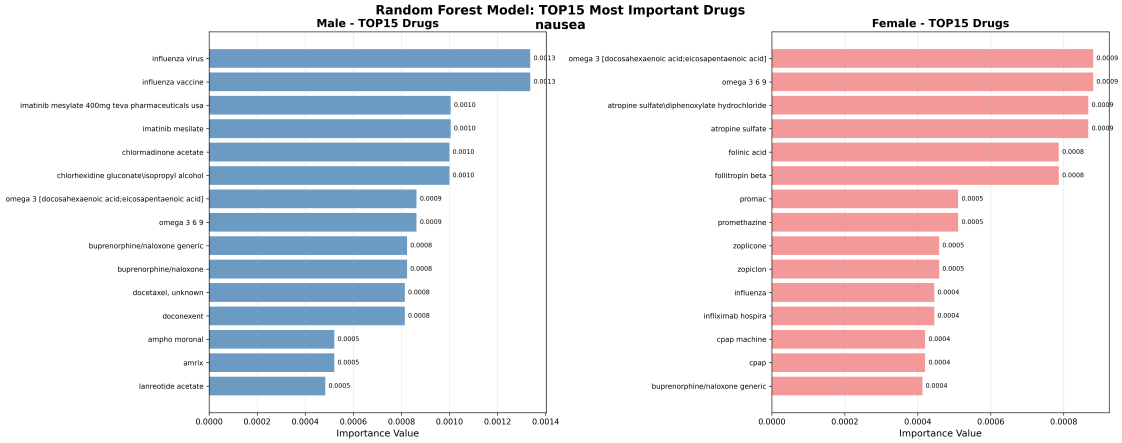

Figure 18: Sex-specific standardized SHAP values for risk drugs in the Random Forest model for "nausea".

Additionally, in the further analysis of sex differences and correlation based on the Random Forest model results for the adverse reaction "nausea". Figure 19 shows a Pearson correlation coefficient of 0.5109 and a Spearman correlation coefficient of 0.5280, indicating a relatively low to moderate correlation. Visualization of the sampled data also reveals considerable differences in normalized importance values between males and females for numerous drugs. For instance, follitropin beta and folinic acid exhibit substantially higher normalized importance in females than in males (difference value = –0.876). Conversely, influenza virus and influenza vaccine show greater feature importance in males (difference value ≈ +0.711). These findings suggest that, although the overall correlation metrics indicate a moderate level of association, the contribution distributions of key risk drug features identified by the Random Forest model for "nausea" are markedly distinct across sex subgroups when examined at the individual drug level.

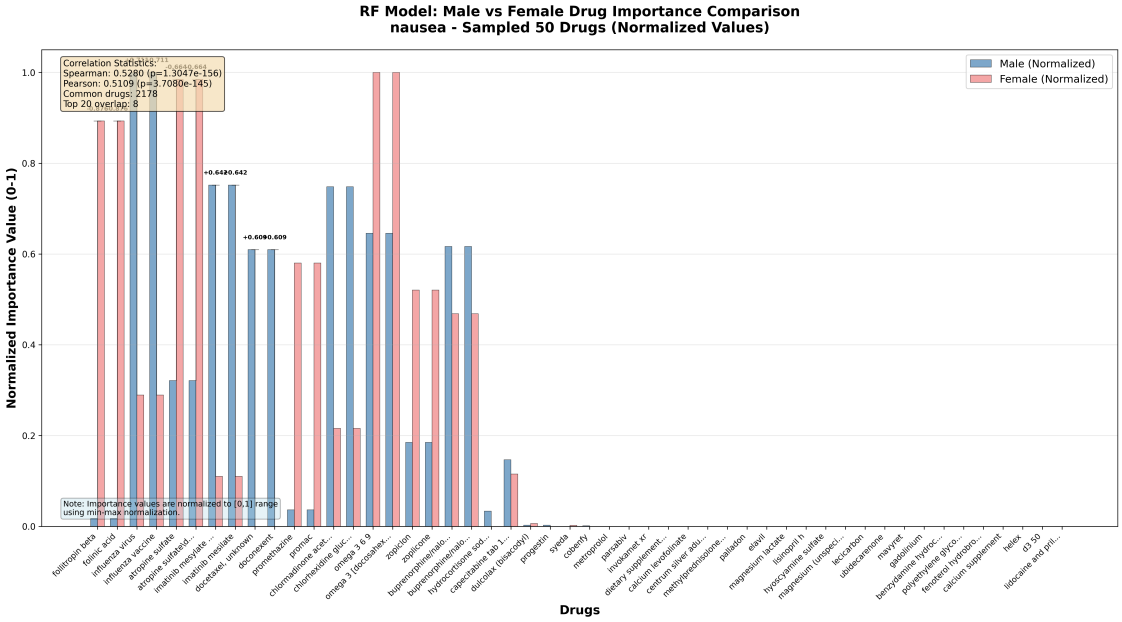

Figure 19: Comparison of sex-specific importance among 50 sampled drugs in the Random Forest model for "nausea"

5.3 Cross-Model Consistency of Sex-Specific Signals

We investigated sex-based differences in adverse drug reactions (ADRs) and identified drugs with sex-specific risk profiles using the ratio of odds ratios (ROR) as the core evaluation metric. The calculation logic proceeds as follows: first, the odds ratio (OR) for ADR occurrence is computed separately for males and females by comparing the exposed group (drug use) with the control group (non-use); the ROR is then derived as the ratio of these two odds ratios:

$$\text{Ratio} = \frac{OR_{male}}{OR_{female}}$$

where OR_{male} denotes the odds of ADR occurrence in males using the drug relative to those not using it, and analogously for females. The specific computational steps are as follows:

1. Model Training: Classification models (Random Forest or LASSO) are constructed and fitted for each specific ADR to obtain the predictive model.

2. Subgroup Stratification: The records for prediction are stratified by sex into male and female subgroups.

3. Probability Prediction: The fitted model is used to predict the probability of ADR occurrence for individuals within both the male and female subgroups.

4. Odds Ratio Estimation: For each subgroup, the mean predicted probabilities are calculated under two counterfactual conditions: presence and absence of the specific drug. The Odds Ratio (OR) for each subgroup is then estimated by calculating the ratio of the mean probability of ADR occurrence with drug exposure to the mean probability without exposure.

Based on reporting data for the top 13 adverse reactions, we applied both linear (LASSO) and nonlinear (Random Forest) modeling approaches. The consistency of the drug–sex interaction signals captured by the two algorithms was visually compared using a scatter plot (Figure 20).

In this plot, the horizontal axis represents the ROR estimated by the LASSO model, while the vertical axis represents the ROR estimated by the Random Forest model, with each point corresponds to the coordinates of the same drug–adverse reaction pair under the two modeling approaches. The results along the x-axis indicate that the risk ratios identified by the LASSO model vary over a wide range (from 0.0 to above 5.0). This suggests that, for certain drugs, the probability of experiencing an ADR exhibits marked asymmetry between males and females: for some drugs the risk is substantially higher in the female subgroup (Ratio > 1.0), whereas others show clear male susceptibility (Ratio < 1.0). Such pronounced numerical fluctuations reveal extensive and non-negligible sex-dependent heterogeneity in drug responses.

In contrast, the Random Forest results displayed along the y-axis are considerably more stable, with most values concentrated near 1.0. This implies that within a nonlinear ensemble framework, the majority of drugs do not present extreme cross-sex risk disparities. Deviations from 1.0 in the Random Forest model occur only when a drug exhibits an exceptionally strong pathogenic signal in a specific sex. This distribution pattern suggests that while sex differences are widespread, the proportion of drugs showing extreme sex-specific risk remains relatively limited. Furthermore, the scatter points do not closely align along the y = x diagonal, indicating a systematic discrepancy in the scale with which the two models identify subgroup differences.

Additionally, we computed and compared the male-to-female drug weight ratios for the top 13 ADRs in both the Relaxed LASSO and Random Forest models, as visualized in Figure 21. In this scatter plot, the x-axis represents the ratio of drug weights between male and female subgroups for a given ADR in the Relaxed LASSO model, while the y-axis represents the corresponding ratio

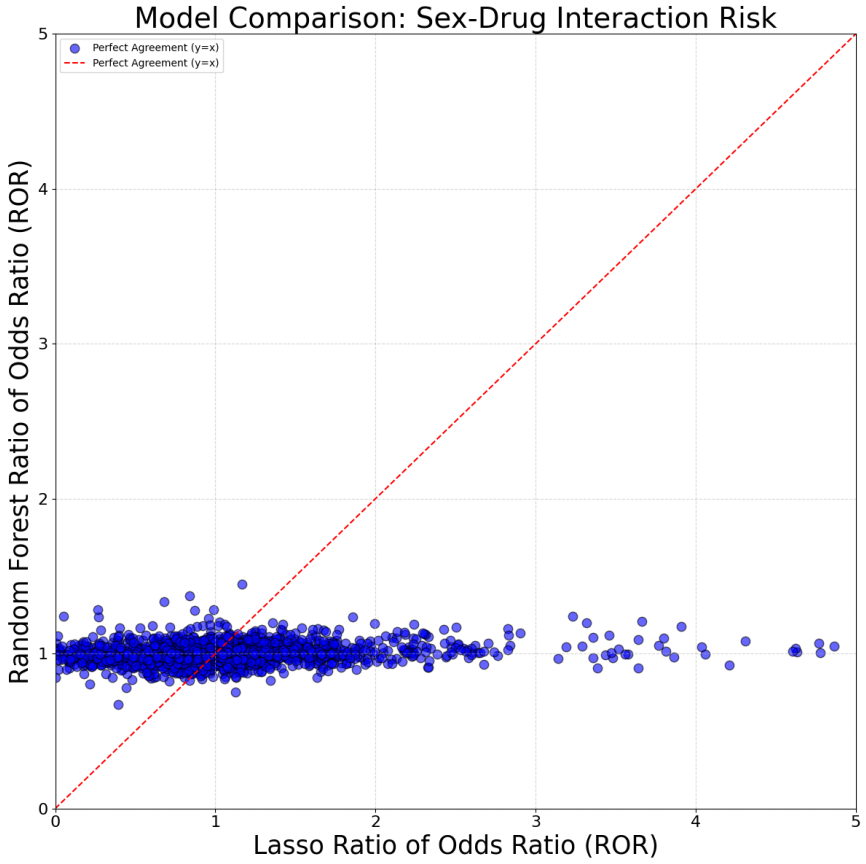

Figure 20: Scatter plot comparing sex-specific ROR estimates from the LASSO and Random Forest models

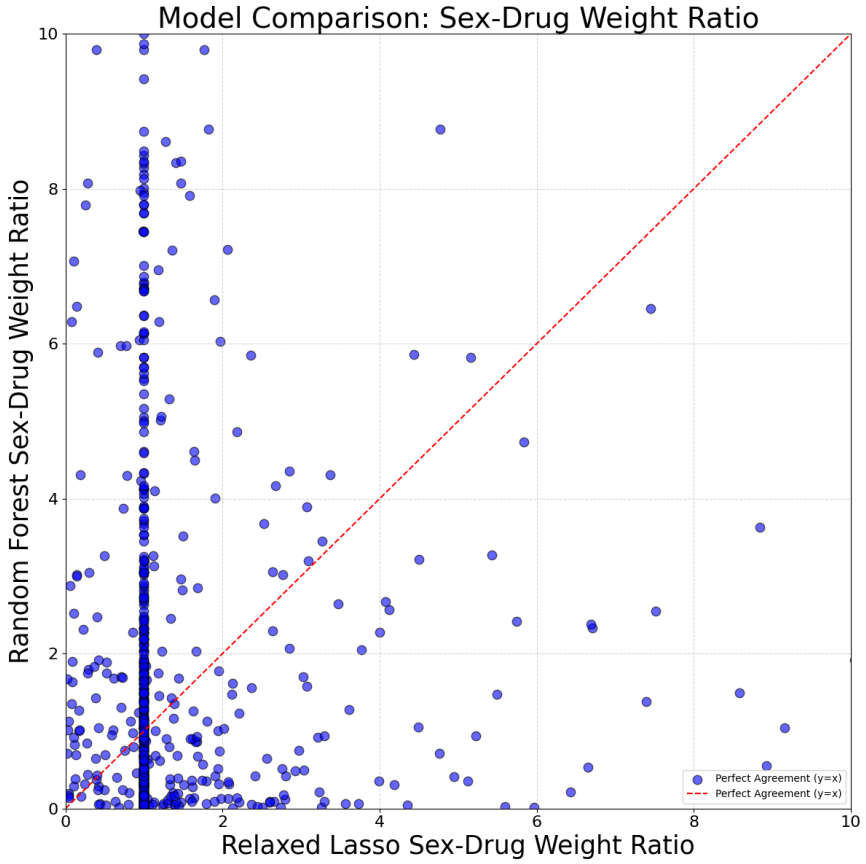

Figure 21: Scatter plot comparing sex-specific drug weight ratios estimated by the Relaxed LASSO and Random Forest models

in the Random Forest model for the same ADR and drug. For both models, many data points deviate significantly from a ratio of 1.0, confirming that the importance of the same drug in adverse reactions differs between male and female subgroups.

However, a clear asymmetry remains in the male-to-female drug weight ratios between the two models, as the scatter points deviate from the $y = x$ line. This indicates that the relative drug effects across sex subgroups are inconsistent between the two frameworks, aligning with the findings presented in Section 5.2. Notably, a considerable number of points in the Relaxed LASSO model are concentrated around $x = 1.0$, whereas the points in the Random Forest model are more widely dispersed. This suggests that the Relaxed LASSO model tends to assign similar effects to many drugs across sexes, while the Random Forest model captures greater variability. This divergence likely stems from the models' varying sensitivity to complex feature interactions and nonlinear relationships.

6 Discussion

This study conducted a sex-stratified modeling analysis of three common adverse drug reactions (ADRs) using FAERS data by applying and comparing two predictive and interpretable machine learning approaches: Relaxed LASSO with interaction terms and Random Forest combined with the SHAP framework. By integrating model performance evaluation with the identification and validation of key risk drugs, this research reveals the complex heterogeneity and commonality of ADR risks across sexes from multiple perspectives. These findings provide empirical evidence supporting the development of a transparent and precise sex-differentiated pharmacovigilance early warning system. The following sections first summarize the primary findings of the study, then provide an in-depth discussion of the sex-specific risk patterns uncovered by each modeling approach and their inherent complementarity, and finally reflects on the significance of multi-method comparisons for both pharmacovigilance research and clinical practice.

6.1 Overall Findings: Method Dependence in Model Performance and Risk Identification

Overall, the two modeling approaches exhibited marked disparities in predictive performance. The Relaxed LASSO consistently demonstrated superior comprehensive predictive efficacy (measured by ACC, ROC-AUC, PR-AUC, and F1 score) across most adverse reactions, particularly excelling in capturing clear linear interaction effects between sex and drugs. In contrast, the Random Forest + SHAP approach offered distinct advantages in interpretability of sex-specific features, albeit with comparatively lower predictive performance.

Notably, the predictive advantage across sex subgroups was not consistent between the two models. For instance, in the case of "fatigue", the Relaxed LASSO model exhibited higher performance in males, whereas the Random Forest model showed only minimal sex differences. This variation in

performance advantage suggests that the sex-specific risk patterns identified and optimized by different modeling approaches are fundamentally distinct [9, 24]. Furthermore, these results imply that a single model may be insufficient to fully capture the multidimensional nature of sex heterogeneity [25, 26].

6.2 Sex-Differentiated Risk Profiles from a Methodological Perspective

Relaxed LASSO directly quantifies the risk disparities of the same drug between males and females by explicitly incorporating sex–drug interaction terms within a linear framework, thereby capturing parametric and linearly additive interaction effects. This method is particularly adept at identifying sex-specific signals that are both statistically strong and sharply contrasting [27, 28]. Consequently, when a sex subgroup contains a substantial number of drugs exhibiting such pronounced effects, the predictive performance—specifically the F1 score—of the Relaxed LASSO model for that subgroup may be notably enhanced [29]. Examples identified in this study include the testosterone agent Striant [30], which poses a prominent risk in males, and the sedative-hypnotic Zolpidem Tartrate ER [31] and the opioid analgesic Morphine Sulfate [32], both of which are more critical in females. These findings not only provide a clinically actionable list of sex-differentiated high-risk drugs but also confirm the utility of interaction term analysis within parametric models for detecting extreme differential signals.

In contrast, the Random Forest model adopts a sex-stratified independent modeling strategy. Compared with the unified modeling approach incorporating interaction terms used in LASSO, this independent strategy offers greater flexibility in fitting the optimal data structure specific to each sex [33]. The results indicate that the Random Forest approach tends to amplify cross-sex performance gaps, suggesting that when models are allowed to learn independently from sex-specific subsets, the potential predictive efficacy revealed may diverge in direction from the interaction effects estimated under global constraints [34]. Moreover, the Random Forest + SHAP approach captures data-driven, prediction-contribution-based overall feature importance. It excels at identifying drug features that—regardless of effect direction—frequently interact with other features and contribute most substantially to improving predictive performance within each sex group [22, 23]. Thus, its performance advantage tends to favor sex subgroups characterized by concentrated feature importance and clearly defined classification patterns.

On one hand, the Random Forest + SHAP analytical pathway, grounded in feature prediction contribution, identifies the drug features that exert the greatest influence on model decisions within each sex group. The results demonstrate that for adverse reactions such as “fatigue” and “nausea”, the Random Forest + SHAP method identified key risk drugs with more pronounced sex differences than those detected by Relaxed LASSO. However, in the case of “death”, certain high-risk drugs consistently ranked as critical predictors for both sexes, reflecting a class of universal risk factors that transcend sex. This underscores a high degree of cross-sex consistency from the perspective of global feature importance, suggesting that some core drug risks carry pan-sex warning significance

[35].

On the other hand, across the analyses of multiple adverse reactions, the Spearman correlation coefficients for the Random Forest model were consistently lower than those for the Relaxed LASSO model. This finding further confirms the Random Forest model's capacity to identify distinct sex-specific risk structures. This trend indicates that as the analysis moves from linear relationships (LASSO) to complex nonlinear interactions (Random Forest), the influence of sex as a demographic variable on drug safety becomes increasingly pronounced. This implies that in clinical pharmacovigilance, relying solely on linear models may underestimate the risk of adverse drug reactions associated with specific drugs in particular sex subgroups.

6.3 Limitations

This study has several limitations. First, during the modeling process, the proportion of positive samples for all adverse reactions was relatively low (5.1%–17.2%), particularly for "fatigue" and "nausea". Although sample balancing techniques were applied to train the models on balanced datasets, caution is warranted when extrapolating model performance to highly imbalanced real-world settings [36]. Future research should explore more advanced imbalanced learning algorithms or adopt cost-sensitive learning frameworks. Additionally, the features used in the current modeling process were restricted to binary drug exposure (yes/no). Subsequent studies could incorporate continuous variables, such as dosage and treatment duration, to further enhance the models' predictive granularity and robustness.

6.4 Future Directions

The two methods—Relaxed LASSO and Random Forest + SHAP—offer complementary insights into the heterogeneity and commonality of drug risks across sexes by analyzing different analytical dimensions: linear interaction effects and global nonlinear feature contributions. Relaxed LASSO provides clear, interpretable sex-comparative coefficients, making it well-suited for hypothesis testing and the generation of explicit sex-differentiated monitoring lists [37]. In contrast, Random Forest + SHAP emphasizes feature contribution predictions within each sex group, facilitating the identification of high-risk drugs involved in complex nonlinear interactions. This multi-method framework not only enhances the robustness of the conclusions but also provides multi-level evidence for understanding why the same drug may exhibit distinct risk-driving mechanisms across sex subgroups.

Through a comprehensive analysis of the sex-specific key risk drug profiles derived from both modeling approaches, we further observed that each method offers distinct insights into the sex heterogeneity of adverse drug reactions. Together, they depict a multi-layered and structurally complex risk landscape. The divergences between the risk profiles identified by the Relaxed LASSO and Random Forest models further suggest that different analytical methods may capture distinct aspects of drug–ADR associations. These findings underscore the value of adopting multi-method,

sex-stratified analytical strategies in drug safety surveillance [38], providing both empirical evidence and methodological references for the future development of more targeted and interpretable personalized medication safety early warning systems.

Future research should aim to integrate physiological, genomic, and broader sex- and gender-related factors [11, 39, 40, 41] to construct more comprehensive sex-differentiated risk prediction models. The continued application of explainable artificial intelligence techniques to subgroup-level risk attribution will help advance precision pharmacovigilance toward greater transparency, fairness, and clinical operability [15, 42, 43]. Ultimately, these methodological advancements are expected to provide clinicians, pharmacists, and regulatory authorities with more targeted decision support, thereby meaningfully improving the safety and efficacy of individualized pharmacotherapy.

7 Conclusion

This study, utilizing large-scale real-world data from the FAERS database, applied and compared two interpretable machine learning frameworks—Relaxed LASSO with interaction terms and Random Forest combined with the SHAP framework—to elucidate the heterogeneity and commonality of adverse drug reaction (ADR) risks across sexes. The results demonstrate significant sex-based disparities in ADR risks, with the two modeling approaches characterizing this complex landscape from complementary perspectives: Relaxed LASSO effectively identifies and quantifies linear sex–drug interaction effects, while the Random Forest + SHAP approach highlights feature contribution within each sex group and uncovers nonlinear risk patterns. This multi-method comparison not only confirms the necessity of sex-stratified analysis but also establishes a methodological foundation for developing interpretable, robust, and clinically applicable pharmacovigilance early warning systems. Looking forward, the integration of multidimensional biomedical data with explainable artificial intelligence (XAI) techniques will further catalyze the advancement of precision pharmacovigilance and personalized medicine.

References

[1] The role of ai in predicting adverse drug reactions: Enhancing patient safety in pharmaceutical practice. *Journal of Population Therapeutics and Clinical Pharmacology*, 31(11):80–89, 2024.

[2] M. Askari et al. Frequency and nature of drug-drug interactions in the intensive care unit. *Pharmacoepidemiology and Drug Safety*, 22(4):430–437, 2013.

[3] R. Raschetti et al. Suspected adverse drug events requiring emergency department visits or hospital admissions. *European Journal of Clinical Pharmacology*, 54(12):959–963, 1999.

[4] D. S. Budnitz et al. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA*, 296(15):1858–1866, 2006.

[5] H. Lee, S. Song, Y. K. Oh, W. Kang, and E. Kim. Is gender still a predisposing factor in contrast-media associated adverse drug reactions? a systematic review and meta-analysis. *European Journal of Radiology*, 89:81–89, 2017.

[6] S. Watson, O. Caster, P. A. Rochon, and H. den Ruijter. Reported adverse drug reactions in women and men: aggregated evidence from globally collected individual case reports during half a century. *EClinicalMedicine*, 17:100188, 2019.

[7] A. Zekarias, S. Watson, S. H. Vidlin, and B. Grundmark. Sex differences in reported adverse drug reactions to covid-19 drugs in a global database of individual case safety reports. *Drug Safety*, 43:1309–1314, 2020.

[8] Emeri Potter, Melissa Reyes, Jennifer Naples, and Gerald Dal Pan. Fda adverse event reporting system (faers) essentials: a guide to understanding, applying, and interpreting adverse event data reported to faers. *Clinical Pharmacology & Therapeutics*, 118(3):567–582, 2025.

[9] Z. Wang, Y. Shi, X. Liu, C. Chen, J. Wen, and R. Wang. Hoddi: A dataset of high-order drug-drug interactions for computational pharmacovigilance. *arXiv preprint arXiv:2502.06274*, 2025.

[10] A. M. Mostafa et al. Innovative tailored semantic embedding and machine learning for precise prediction of drug-drug interaction seriousness. *IEEE Access*, 13:49249–49270, 2025.

[11] K. S. Just et al. Adverse drug reactions in the emergency department: Is there a role for pharmacogenomic profiles at risk?—results from the adred study. *Journal of Clinical Medicine*, 9(6):1801, 2020.

[12] R. Wakao, I. M. Lönnstedt, Y. Aoki, and R. E. Chandler. The use of subgroup disproportionality analyses to explore the sensitivity of a global database of individual case safety reports to known pharmacogenomic risk variants common in japan. *Drug Safety*, 44(6):681–697, 2021.

[13] P. Alberti and G. Cavaletti. Management of side effects in the personalized medicine era: chemotherapy-induced peripheral neuropathy. In *Pharmacogenomics Drug Discovery and Development*, pages 301–322. 2014.

[14] M. J. Alomar. Factors affecting the development of adverse drug reactions. *Saudi Pharmaceutical Journal*, 22:83–94, 2014.

[15] T. H. Vo, N. T. K. Nguyen, Q. H. Kha, and N. Q. K. Le. On the road to explainable ai in drug-drug interactions prediction: A systematic review. *Computational and Structural Biotechnology Journal*, 20:2112–2123, 2022.

[16] Sunwoo Jung and Sunyong Yoo. Interpretable prediction of drug-drug interactions via text embedding in biomedical literature. *Computers in Biology and Medicine*, 185:109496, 2025.

[17] Zijiao Liu and Ruolan Chen. Drug-related demyelination: A retrospective pharmacovigilance study based on the faers database. *Advances in Clinical Medicine*, 15(9):1703–1715, 2025.

[18] Guillermo Rodriguez-Nava, Goar Egoryan, Daniela Patricia Trelles-Garcia, Maria Adriana Yanez-Bello, and Ricardo Murguia-Fuentes. Disproportionality analysis of anaphylactic reactions after vaccination with messenger rna coronavirus disease 2019 vaccines in the united states. *Annals of Allergy, Asthma & Immunology*, 127(1):139–140, 2021.

[19] Nicolai Meinshausen. Relaxed lasso. *Computational Statistics & Data Analysis*, 52(1):374–393, 2007.

[20] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[21] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

[22] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[23] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[24] Jeffrey Sun, Cheuk-Kay Sun, Yun-Xuan Tang, Tzu-Chi Liu, and Chi-Jie Lu. Application of shap for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification. In *Healthcare*, volume 11, page 2000. MDPI, 2023.

[25] Yuting Shan, Lee Cheung, Yuqi Zhou, Yingbo Huang, and R Stephanie Huang. A systematic review on sex differences in adverse drug reactions related to psychotropic, cardiovascular, and analgesic medications. *Frontiers in Pharmacology*, 14:1096366, 2023.

[26] Flavia Franconi and Ilaria Campesi. Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women. *British journal of pharmacology*, 171(3):580–594, 2014.

[27] Carl M O'Brien. Statistical learning with sparsity: the lasso and generalizations. 2016.

[28] Offie P Soldin and Donald R Mattison. Sex differences in pharmacokinetics and pharmacodynamics. *Clinical pharmacokinetics*, 48(3):143–157, 2009.

[29] Evangelia Christodoulou, Jie Ma, Gary S Collins, Ewout W Steyerberg, Jan Y Verbakel, and Ben Van Calster. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, 110:12–22, 2019.

[30] Rebecca Vigen, Colin I O'Donnell, Anna E Barón, Gary K Grunwald, Thomas M Maddox, Steven M Bradley, Al Barqawi, Glenn Woning, Margaret E Wierman, Mary E Plomondon, et al. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. *Jama*, 310(17):1829–1836, 2013.

[31] US Food, Drug Administration, et al. Risk of next-morning impairment after use of insomnia drugs; fda requires lower recommended doses for certain drugs containing zolpidem (ambien, ambien cr, edluar, and zolpimist). *Silver Spring, MD: Drug Safety Communications*, 2013.

[32] Guilherme S Lopes, Suzette Bielinski, Ann M Moyer, Debra J Jacobson, Liwei Wang, Ruoxiang Jiang, Nicholas B Larson, Virginia M Miller, Ye Zhu, Dana C Cavanaugh, et al. Sex differences in type and occurrence of adverse reactions to opioid analgesics: a retrospective cohort study. *BMJ open*, 11(6):e044157, 2021.

[33] Rui Wang, Stephen W Lagakos, James H Ware, David J Hunter, and Jeffrey M Drazen. Statistics in medicine—reporting of subgroup analyses in clinical trials. *New England Journal of Medicine*, 357(21):2189–2194, 2007.

[34] Todd A Alonzo. Clinical prediction models: a practical approach to development, validation, and updating: by ewout w. steyerberg, 2009.

[35] HT Rakshith, S Lohita, Anvil Preem Rebello, Prakash S Goudanavar, and N Raghavendra Naveen. Sex differences in drug effects and/or toxicity in oncology. *Current Research in Pharmacology and Drug Discovery*, 4:100152, 2023.

[36] Bilal Mirza, Wei Wang, Jie Wang, Howard Choi, Neo Christopher Chung, and Peipei Ping. Machine learning and integrative analysis of biomedical big data. *Genes*, 10(2):87, 2019.

[37] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[38] Qiaozhi Hu, Yuxian Chen, Dan Zou, Zhiyao He, and Ting Xu. Predicting adverse drug event using machine learning based on electronic health records: a systematic review and meta-analysis. *Frontiers in Pharmacology*, 15:1497397, 2024.

[39] Andreea C Brabete, Lorraine Greaves, Mira Maximos, Ella Huber, Alice Li, and Mê-Linh Lê. A sex-and gender-based analysis of adverse drug reactions: a scoping review of pharmacovigilance databases. *Pharmaceuticals*, 15(3):298, 2022.

[40] Franck Mauvais-Jarvis, Noel Bairey Merz, Peter J Barnes, Roberta D Brinton, Juan-Jesus Carrero, Dawn L DeMeo, Geert J De Vries, C Neill Epperson, Ramaswamy Govindan, Sabra L Klein, et al. Sex and gender: modifiers of health, disease, and medicine. *The Lancet*, 396(10250):565–582, 2020.

[41] Ann M Moyer, Eric T Matey, and Virginia M Miller. Individualized medicine: Sex, hormones, genetics, and adverse drug reactions. *Pharmacology research & perspectives*, 7(6):e00541, 2019.

[42] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I Madai, and Precise4Q Consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making*, 20(1):310, 2020.

[43] US FDA. Diversity action plans to improve enrollment of participants from underrepresented populations in clinical studies; draft guidance for industry. *DHHS, ed. Vol Draft Guidance2024*, 2024.

A Appendix

A.1 Sensitivity Analysis of the Impact of PRR Threshold Variation on Model Prediction Robustness

A.1.1 Research Purpose

This experiment evaluates the sensitivity of downstream machine learning model performance—specifically utilizing Relaxed LASSO—to variations in the Proportional Reporting Ratio (PRR) threshold, a fundamental parameter in disproportionality analysis. By comparing model outcomes across different threshold conditions, we investigate how the number of identified drugs and the accuracy of classification predictions are impacted, thereby assessing the robustness of the model under varying signal selection criteria.

A.1.2 Methodology

1. PRR Thresholding: Three distinct PRR thresholds (1.5, 2.0, and 2.5) were established as inclusion criteria for drug selection prior to model entry.

2. Model Construction: For the three common adverse reactions ("death", "fatigue", and "nausea"), the Relaxed LASSO model was used to perform binary classification predictions on male and female subsets independently.

3. Performance Metrics: Model performance under varying threshold conditions was quantified using four core evaluation metrics: accuracy (ACC), ROC-AUC, PR-AUC, and F1 score.

Table 6: Signal counts under different thresholds for disproportionality analysis

Min a	Min PRR	Min Chi2	Signal Count	Delta(vs Baseline)
3	1.5	4	1825	+603
3	1.5	6.63	1721	+499
3	2	4	1222	0
3	2	6.63	1222	0
3	2.5	4	882	-340
3	2.5	6.63	882	-340
4	1.5	4	1825	+603
4	1.5	6.63	1721	+499
4	2	4	1222	0
4	2	6.63	1222	0
4	2.5	4	882	-340
4	2.5	6.63	882	-340

Table 7: Sensitivity analysis of PRR thresholds under balanced sampling based on the Relaxed LASSO model

Adverse Reaction Type	PRR Treshold	Male				Female			
		ACC	ROC-AUC	PR-AUC	F1	ACC	ROC-AUC	PR-AUC	F1
death	PRR=1.5	0.7156	0.7793	0.8073	0.7435	0.7233	0.7873	0.7669	0.6851
	PRR=2	0.6956	0.7719	0.7975	0.7247	0.7013	0.7737	0.7596	0.6663
	PRR=2.5	0.6886	0.7580	0.7754	0.7174	0.7024	0.7675	0.7561	0.6421
fatigue	PRR=1.5	0.6777	0.7193	0.6663	0.6296	0.5976	0.6391	0.6486	0.5816
	PRR=2	0.6634	0.7094	0.6942	0.6715	0.6113	0.6590	0.6640	0.5679
	PRR=2.5	0.6480	0.6977	0.7033	0.6779	0.6318	0.6767	0.6561	0.5813
nausea	PRR=1.5	0.6409	0.6569	0.5679	0.4983	0.6394	0.6756	0.6946	0.6816
	PRR=2	0.6369	0.6658	0.5972	0.5089	0.6072	0.6413	0.6531	0.6704
	PRR=2.5	0.6398	0.6823	0.5954	0.5465	0.6094	0.6498	0.6586	0.6318

A.1.3 Results Analysis

The PRR thresholds were adjusted around the Evans Criteria ($a = 3$, $PRR = 2$, $\chi^2 = 4$), with the corresponding experimental data presented in Table 6. The results indicate that the PRR threshold has a substantial impact on the number of signals identified. Based on the data shown in Table 7, several key conclusions can be drawn:

1. Model performance fluctuates with increasing PRR threshold: Model performance is significantly sensitive to changes in the PRR threshold. For example, in the classification of "nausea", as the PRR threshold increased from 1.5 to 2.5, the ROC-AUC for males increased from 0.6569 to 0.6823. This suggests that a higher threshold filters out more background noise, allowing the model to learn from more strongly associated signals and thereby improving predictive accuracy.

2. Sex-based disparities in performance: For most ADR types, predictive metrics were slightly higher for male data than for female data. For instance, at a threshold of PRR = 1.5 for the death" outcome, the PR-AUC for males was 0.8073, compared to 0.7669 for females.

3. Model consistency: Although the number of identified drugs decreases as the threshold increases, the predictive performance of the Relaxed LASSO model remained relatively stable across key metrics, confirming its adaptability in handling signals of varying strength.

In summary, the sensitivity analysis confirms that both the number of identified drugs and the classification performance of the model are highly sensitive to the choice of PRR threshold. Although a higher PRR threshold generally leads to improvements in evaluation metrics such as ROC-AUC, this occurs at the expense of reduced drug coverage. Therefore, selecting a specific threshold (e.g., PRR = 2) as a benchmark to balances signal strength and coverage is both reasonable and statistically justified for subsequent analyses.

Table 8: Performance comparison of the Relaxed LASSO model under balanced sampling and threshold-based methods

Adverse Reaction Type	Modeling Method	Male				Female			
		ACC	ROC-AUC	PR-AUC	F1	ACC	ROC-AUC	PR-AUC	F1
death	Relaxed LASSO	0.6956	0.7719	0.7975	0.7247	0.7013	0.7737	0.7596	0.6663
	Relaxed LASSO-threshold	0.6844	0.7658	0.8017	0.7425	0.6686	0.7660	0.7371	0.6967
fatigue	Relaxed LASSO	0.6634	0.7094	0.6942	0.6715	0.6113	0.6590	0.6640	0.5679
	Relaxed LASSO-threshold	0.6593	0.7103	0.7144	0.6326	0.6140	0.6659	0.6677	0.5035
nausea	Relaxed LASSO	0.6369	0.6658	0.5972	0.5089	0.6072	0.6413	0.6531	0.6704
	Relaxed LASSO-threshold	0.6230	0.6602	0.5836	0.4671	0.5967	0.6440	0.6765	0.6532

A.2 Comparative Experiment on Modeling Strategies

In Table 8, "Relaxed LASSO" refers to training the Relaxed LASSO model using balanced sampling, while "Relaxed LASSO-threshold" denotes training on the full sample and subsequently adjusting the classification threshold for prediction. The experimental results comparing the performance of Relaxed LASSO across these sampling strategies indicate that both approaches yield reliable modeling performance in identifying sex-specific adverse drug reaction risks, although distinct strategic differences emerge across various evaluation metrics.

From the perspective of overall performance, both strategies achieved largely consistent ROC-AUC values across most adverse reaction types (male: 0.6602–0.8094; female: 0.6413–0.8228), indicating that the model's ability to rank positive and negative samples remains stable regardless of the sampling strategy. This finding confirms the robustness of the Relaxed LASSO approach in handling high-dimensional, sparse ADR data and demonstrates that its interaction term design and two-stage regularization mechanism effectively capture sex–drug interaction effects.

In terms of the precision–recall trade-off, the full-sample strategy (Relaxed LASSO-threshold) achieved superior PR-AUC values for severe adverse reactions such as "death" and "fatigue" (male "death": 0.8017 vs. 0.7975; male "fatigue": 0.7144 vs. 0.6942), suggesting improved precision in identifying positive samples within imbalanced data environments. This advantage likely stems from the retention of the original data distribution in the full sample, which enables the model to more accurately learn the feature patterns of the minority (positive) class.

However, with respect to classification balance, the balanced sampling strategy demonstrated more stable F1 scores across most scenarios, particularly in the female subgroup (e.g., "nausea": 0.6704 vs. 0.6532), indicating a more balanced trade-off between precision and recall. This divergence reflects the fundamental distinction between the two strategies: balanced sampling optimizes the decision boundary by artificially equalizing class distributions, whereas the full-sample strategy more faithfully reflects the imbalanced structure of real-world reporting data.

Table 9: Performance comparison of the Random Forest model under balanced sampling and weighting methods

Adverse Reaction Type	Modeling Method	Male				Female			
		ACC	ROC-AUC	PR-AUC	F1	ACC	ROC-AUC	PR-AUC	F1
Death	RF+SHAP	0.6655	0.7421	0.7536	0.5832	0.6750	0.7532	0.7563	0.6353
	RF+SHAP-weight	0.6667	0.7428	0.7613	0.5817	0.6694	0.7534	0.7655	0.6202
Fatigue	RF+SHAP	0.6362	0.6713	0.6583	0.6114	0.5900	0.6430	0.6410	0.5053
	RF+SHAP-weight	0.6466	0.7055	0.7007	0.5957	0.5956	0.6478	0.6542	0.4743
Nausea	RF+SHAP	0.5983	0.6329	0.6171	0.5393	0.5830	0.6351	0.6369	0.4441
	RF+SHAP-weight	0.6146	0.6535	0.6592	0.5376	0.5700	0.6288	0.6239	0.4132

In comparing the experimental results of the "RF+SHAP" (balanced sampling) and "RF+SHAP-weight" (full sample with class weight adjustment) modeling strategies, as presented in Table 9, this study observed distinct performance differences and patterns between the two approaches. Overall, both methods exhibited comparable performance in terms of ROC-AUC and PR-AUC, indicating that the model's ability to rank and distinguish positive from negative samples remains relatively stable and is less affected by the sampling strategy. However, regarding accuracy (ACC) and F1 score—particularly for specific adverse reaction types—the balanced sampling approach (RF+SHAP) tended to demonstrate superior or equivalent comprehensive classification performance. From a sex-specific perspective, RF+SHAP performed better on key metrics for "fatigue" in male data, while maintaining advantages in female data for reactions such as "death".

These findings further confirm that balanced sampling mitigates class imbalance caused by differences in reporting frequencies across sexes, enabling the model to learn risk patterns more equitably within both male and female subgroups. In contrast, while the RF+SHAP-weight method attempted to capture information from the full dataset through weight adjustment, it was slightly inferior to the balanced sampling strategy in overall classification performance, particularly in terms of F1 score, which reflects the trade-off between precision and recall under the task settings and data distribution examined in this experiment.

In summary, this experiment demonstrates that when constructing sex-specific predictive and interpretable models for ADRs, adopting a balanced sampling strategy (RF+SHAP) can enhance the model's ability to identify positive adverse reaction cases and improve overall classification harmony without significantly compromising discriminative capacity (as measured by AUC). Consequently, this approach is more conducive to subsequent SHAP-based interpretation of sex-differentiated drug risk profiles.