# Machine Learning Models for Predicting Serious Adverse Event Reports in GLP-1 Receptor Agonists Using FAERS, 2015-2025

CHAPMAN UNIVERSITY
SCHOOL OF PHARMACY

RWD148

Zhouzhou Chu, MSc[1], Daniel O. Umoru, BPharm[1], Ang Li, PhD[2], Ariel Tran, PharmD[1], Bowei Tian, BSc[2], Sherry Yun WANG, PhD[1] (yunwang@chapman.edu)
[1] School of Pharmacy, Chapman University, Irvine, CA, USA
[2]Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA

## BACKGROUND

- GLP-1 receptor agonists (GLP-1RAs) increasingly prescribed for type 2 diabetes and obesity, with prescriptions rising 587% for weight-related indications (2019–2024) [1]; ~12% of U.S. adults report ever use, rising to 43% among those with diabetes [2].
- This adoption drives a surge in FDA Adverse Event Reporting System (FAERS) Database safety reports: >187,000 GI adverse events (2007–2023) [3] and 6,751 acute pancreatitis reports (2005–2023, 98.3% classified as serious [4]).
- Standard FAERS pharmacovigilance relies on disproportionality methods (e.g., ROR, PRR) [5]
- ML models have demonstrated strong FAERS-based serious AE prediction for oncology patients [6], yet no prior study has applied ML-based seriousness classification to GLP-1RAs at scale.

## OBJECTIVE

- To develop and compare logistic regression and random forest models for predicting serious outcome classification in GLP-1RA-associated FAERS reports (2015–2025)
- To identify the clinical, demographic, and reporting-context features associated with seriousness classification across branded and generic GLP-1RA products.

## METHODS

- **Data source:** FAERS Q1 2015–Q1 2025; free-text drug names mapped to RxNorm via NLM RxNav API (score ≥9.0) with ingredient-level lifting; 242,312 case-molecule observations; 26.3% classified serious.
- **Features:** Age, sex, reporter type, indication bucket, molecule identity, brand/generic/mixed exposure, and 12 binary AE category indicators from 6,752 MedDRA Preferred Terms.
- **Models:** Logistic regression (L2) and random forest (500 trees), both with balanced class weights; group-aware 75/25 split; post-hoc Platt and isotonic calibration; evaluated by ROC-AUC, PR-AUC, and Brier score.

## RESULTS

- **Model performance:** LR ROC-AUC = 0.876; RF ROC-AUC = 0.916; isotonic calibration reduced Brier score from 0.143→0.118 (LR) and 0.122→0.096 (RF); RF + Isotonic was the best-calibrated model (Figures 1–2).
- **LR SHAP (Figure 3):** Reporter = Consumer was the most influential predictor (|SHAP| = 0.72), followed by Other AE Category (0.51), Injection-Site Reactions (0.46), and Dosing/Administration Issues (0.42).
- **RF SHAP (Figure 4):** Dosing/Administration Issues (0.064) and Injection-Site Reactions (0.063) were the top predictors; their presence shifted predictions toward non-serious outcomes; molecule identity (Semaglutide = 0.04, Tirzepatide = 0.037) followed.

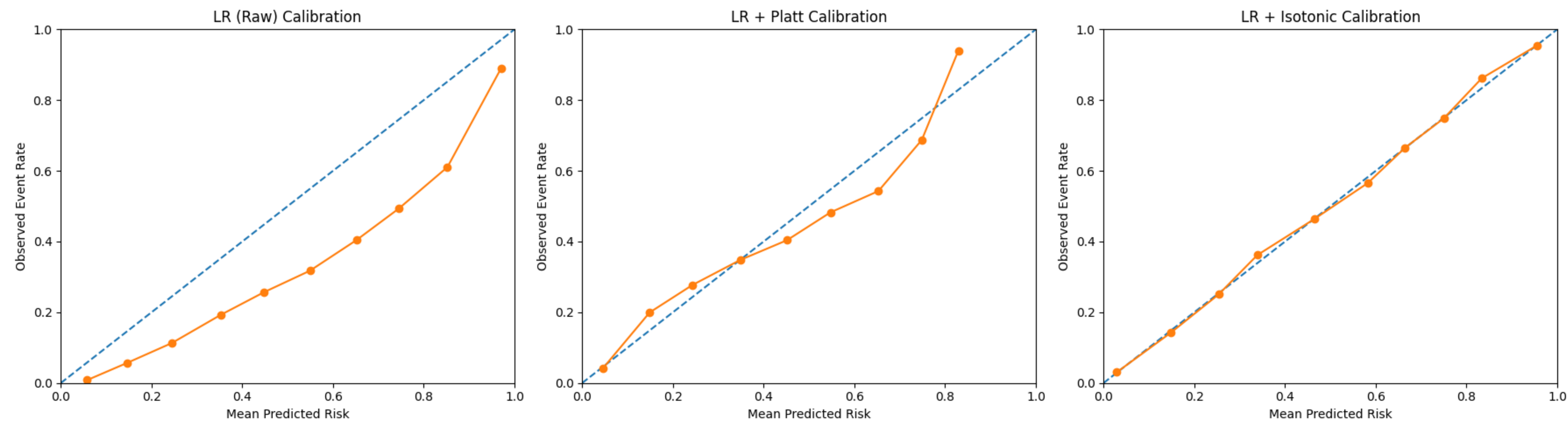


**Figure 1. Calibration Curves: Logistic Regression (Raw, Platt, and Isotonic)**

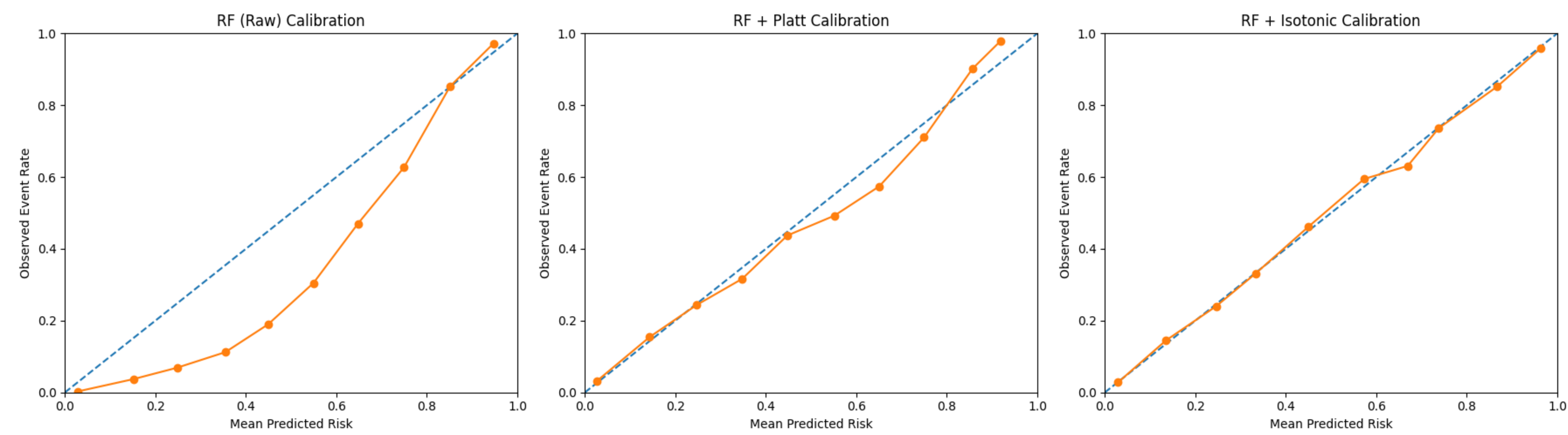


**Figure 2. Calibration Curves: Random Forest (Raw, Platt, and Isotonic)**

## DISCUSSION

- Seriousness within FAERS is more strongly associated with reporting context and adverse event phenotype than by drug identity alone; Reporter = Consumer likely reflects structural reporting behavior rather than true clinical severity.
- Gastrointestinal disorders ranked consistently low in SHAP importance across both models despite being the second most frequently reported AE category.

## LIMITATION

- FAERS reports are subject to reporting bias, underreporting, and absence of exposure denominators; results reflect reporting patterns rather than true incidence.
- Seriousness labels based on outcome codes may be incomplete; SHAP values reflect model associations, not causal relationships.

## CONCLUSION

- Both LR (ROC-AUC 0.876) and RF (ROC-AUC 0.916) effectively classified serious GLP-1RA adverse event reports in FAERS, with strong calibration after isotonic adjustment.
- Seriousness classification was more closely linked to adverse event phenotype and reporting context rather than molecule identity, with dosing/administration issues, injection-site reactions, and reporter type as the most consistent predictors.
- These results support calibrated, interpretable ML as a scalable approach to GLP-1RA pharmacovigilance using routinely collected real-world data.

## REFERENCES

[1] FAIR Health. *Obesity and GLP-1 Drugs: A Claims-Based Analysis* [white paper]. FAIR Health; 2025.

[2] Kirzinger A, Montero A, Sparks G, Hamel L, Brodie M. KFF Health Tracking Poll – May 2024: The Public's Use and Views of GLP-1 Drugs. Kaiser Family Foundation; 2024.

[3] Osei SP, Akomaning E, Florut TF, et al. Gastrointestinal safety assessment of GLP-1 receptor agonists in the US: a real-world adverse events analysis from the FAERS database. *Diagnostics*. 2024;14(24):2829. doi:10.3390/diagnostics14242829

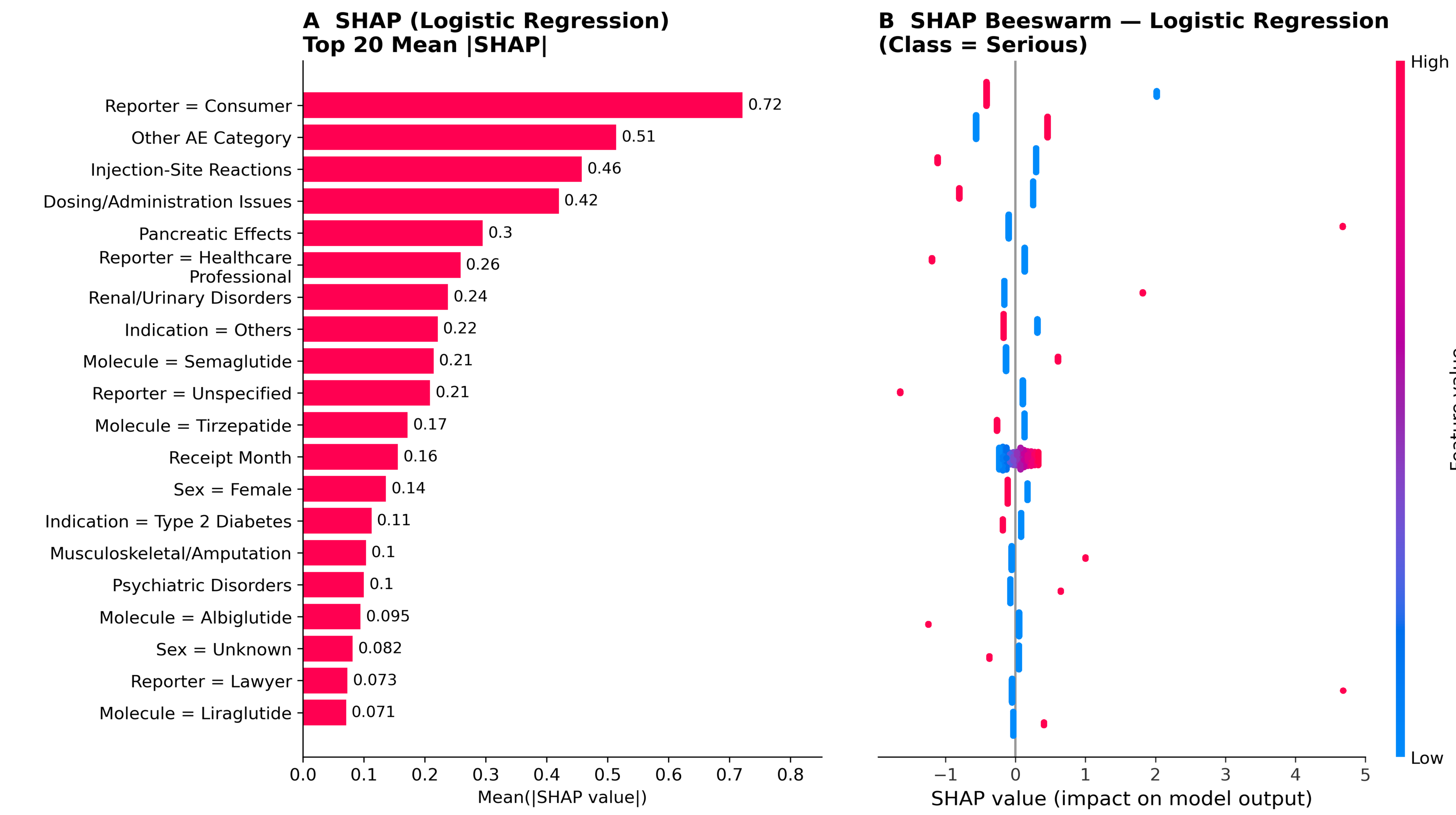


**Figure 3. SHAP Feature Importance: Logistic Regression Model**

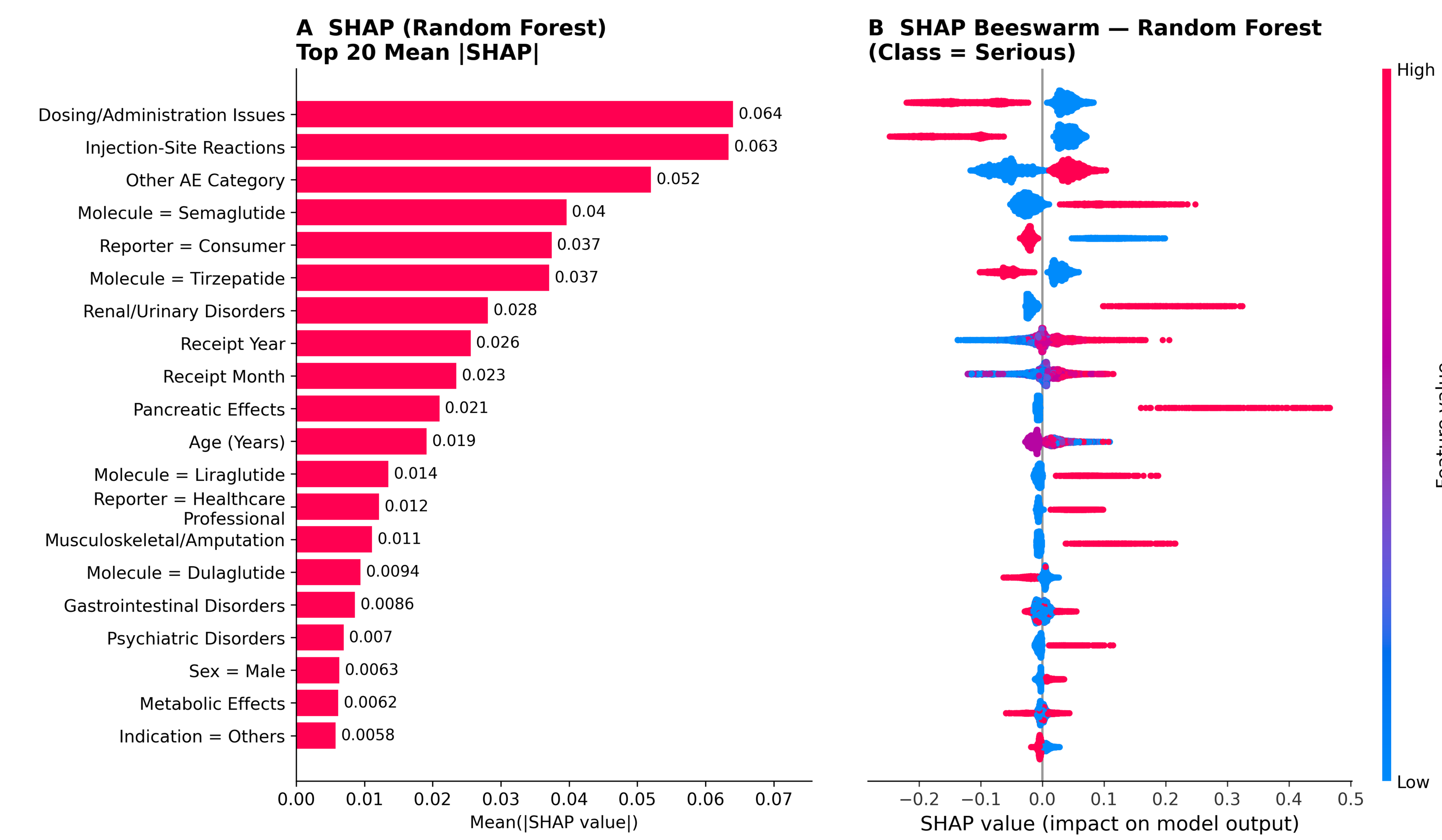


**Figure 4. SHAP Feature Importance: Random Forest Model**

## REFERENCES

[4] Guo H, Guo Q, Li Z, Wang Z. Association between different GLP-1 receptor agonists and acute pancreatitis: case series and real-world pharmacovigilance analysis. *Front Pharmacol*. 2024;15:1461398. doi:10.3389/fphar.2024.1461398

[5] Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA Adverse Event Reporting System. *Int J Med Sci*. 2013;10(7):796-803. doi:10.7150/ijms.6048

[6] Zhao LX, Wang C, Zou J, Ranasinghe P. Machine learning model for predicting severe adverse events in oncology patients using the FDA Adverse Event Reporting System. *JCO Clin Cancer Inform*. 2026;10:e2500081. doi:10.1200/CCI-25-00081

*Presented at ISPOR Conference, Pennsylvania, Philadelphia May 20, 2026*