JIBAS

Journal of Insights in Basic and Applied Sciences

Volume 1, Issue 1, 2025, Pages: 27-34

Journal homepage: https://jibas.com.ly/index.php/jibas/en

Pharmacovigilance and Adverse Drug Reaction Patterns: A Data-Driven Study on Drug Safety in Clinical Practice

Nabil Saad Miftah *
Higher Institute of Medical Sciences, Bani Waleed, Libya

اليقظة الدوائية وأنماط الآثار الجانبية للأدوية: دراسة قائمة على البيانات حول سلامة الأدوية في الممارسة السريرية

نبيل سعد مفتاح*
المعهد العالي للعلوم الطبية، بني وليد، ليبيا

*Corresponding author: nabilmohamed22379@gmail.com

Received: April 11, 2025	Accepted: July 08, 2025	Published: July 17, 2025

Abstract:
Pharmacovigilance is the systematic process of monitoring and analyzing adverse drug reactions (ADRs) to ensure patient safety after medications reach the market. This paper reviews contemporary data-driven approaches to pharmacovigilance, focusing on how large healthcare databases and novel analytics can reveal ADR patterns. We describe traditional data sources such as spontaneous reporting systems (e.g. the WHO's VigiBase and FDA's FAERS) and electronic health records (EHRs), as well as emerging real-world data (RWD) sources like social media and patient registries. Key methods including disproportionality statistics and machine learning are discussed, with examples of recent studies using open ADR datasets. We highlight experiments using publicly available data, such as mining FAERS reports with disproportionality analysis and developing ADR reference sets for validation. For instance, one systematic review notes that in the U.S., ADRs accounted for ~6% of hospital admissions in 2011 and cost billions. We also examine regional patterns (e.g. differences in ADR reporting between Japan and global data) and discuss challenges like underreporting and data quality. Figures and tables illustrate the pharmacovigilance process and ADR data characteristics. By leveraging modern data science, stakeholders can detect safety signals earlier and tailor drug-safety practices to populations worldwide.

Keywords: Pharmacovigilance; adverse drug reaction; data-driven analysis; real-world data; drug safety; signal detection.

الملخص:
اليقظة الدوائية هي عملية منهجية لرصد وتحليل التفاعلات الدوائية الضارة (ADRs) لضمان سلامة المرضى بعد طرح الأدوية في السوق. تستعرض هذه الورقة البحثية مناهج اليقظة الدوائية المعاصرة القائمة على البيانات، مع التركيز على كيفية كشف قواعد بيانات الرعاية الصحية الضخمة والتحليلات الحديثة عن أنماط التفاعلات الدوائية الضارة. نصف مصادر البيانات التقليدية، مثل أنظمة الإبلاغ التلقائي) مثل VigiBase التابع لمنظمة الصحة العالمية ونظام FAERS التابع لإدارة الغذاء والدواء الأمريكية (والسجلات الصحية الإلكترونية(EHRs) ، بالإضافة إلى مصادر البيانات الواقعية الناشئة، مثل منصات التواصل الاجتماعي وسجلات المرضى. نناقش أساليب رئيسية، بما في ذلك إحصاءات عدم التناسب والتعلم الآلي، مع أمثلة لدراسات حديثة استخدمت مجموعات بيانات التفاعلات الدوائية الضارة المفتوحة. نسلط الضوء على التجارب التي استخدمت بيانات متاحة للعامة، مثل تحليل تقارير FAERS باستخدام تحليل عدم التناسب، وتطوير مجموعات مرجعية للتفاعلات الدوائية الضارة للتحقق من صحتها. على سبيل المثال، تشير إحدى المراجعات المنهجية إلى أن التفاعلات الدوائية الضارة في الولايات المتحدة شكلت حوالي 6% من حالات دخول المستشفيات عام 2011، وبلغت تكلفتها مليارات الدولارات. ندرس أيضًا الأنماط الإقليمية (مثل الاختلافات في الإبلاغ عن الآثار الضارة للأدوية بين اليابان والبيانات العالمية) ونناقش تحديات مثل نقص الإبلاغ وجودة البيانات. توضح الأشكال والجداول عملية اليقظة الدوائية وخصائص بيانات الآثار الضارة للأدوية. من خلال الاستفادة من علوم البيانات الحديثة، يمكن للجهات المعنية اكتشاف مؤشرات السلامة مبكرًا وتصميم ممارسات سلامة الأدوية بما يتناسب مع السكان حول العالم.

الكلمات المفتاحية: اليقظة الدوائية؛ الآثار الضارة للأدوية؛ التحليل القائم على البيانات؛ بيانات واقعية؛ سلامة الأدوية؛ كشف المؤشرات.

Introduction

Pharmacovigilance (PV) is the science of detecting, assessing, and preventing adverse drug reactions (ADRs) after medicines are marketed. In practice, this means collecting reports of suspected ADRs, analyzing the data for safety signals, and taking action when needed. ADRs are a major public health issue: one review estimated that in the United States ADRs were responsible for almost 6% of hospitalizations in 2011 and cost the healthcare system billions of dollars. Another study notes that ADRs are a leading cause of death in the U.S., causing more fatalities than diseases like lung disease or diabetes (Shin et al., 2021). Thus, ongoing monitoring of drug safety is essential.

Modern pharmacovigilance increasingly relies on data-driven methods. Vast amounts of data are available from global and national reporting systems, as well as from electronic health records (EHRs), insurance claims, and even patient narratives. For example, the World Health Organization's global database (VigiBase) already contains over 35 million individual case safety reports (ICSRs) as of 2023 (World Health Organization., 2023). This rich data allows sophisticated analyses to identify ADR patterns. Figure 1 below illustrates a simplified PV workflow, from data collection through analysis and signal detection. The workflow is labor-intensive, which motivates the use of machine learning and other automated tools.

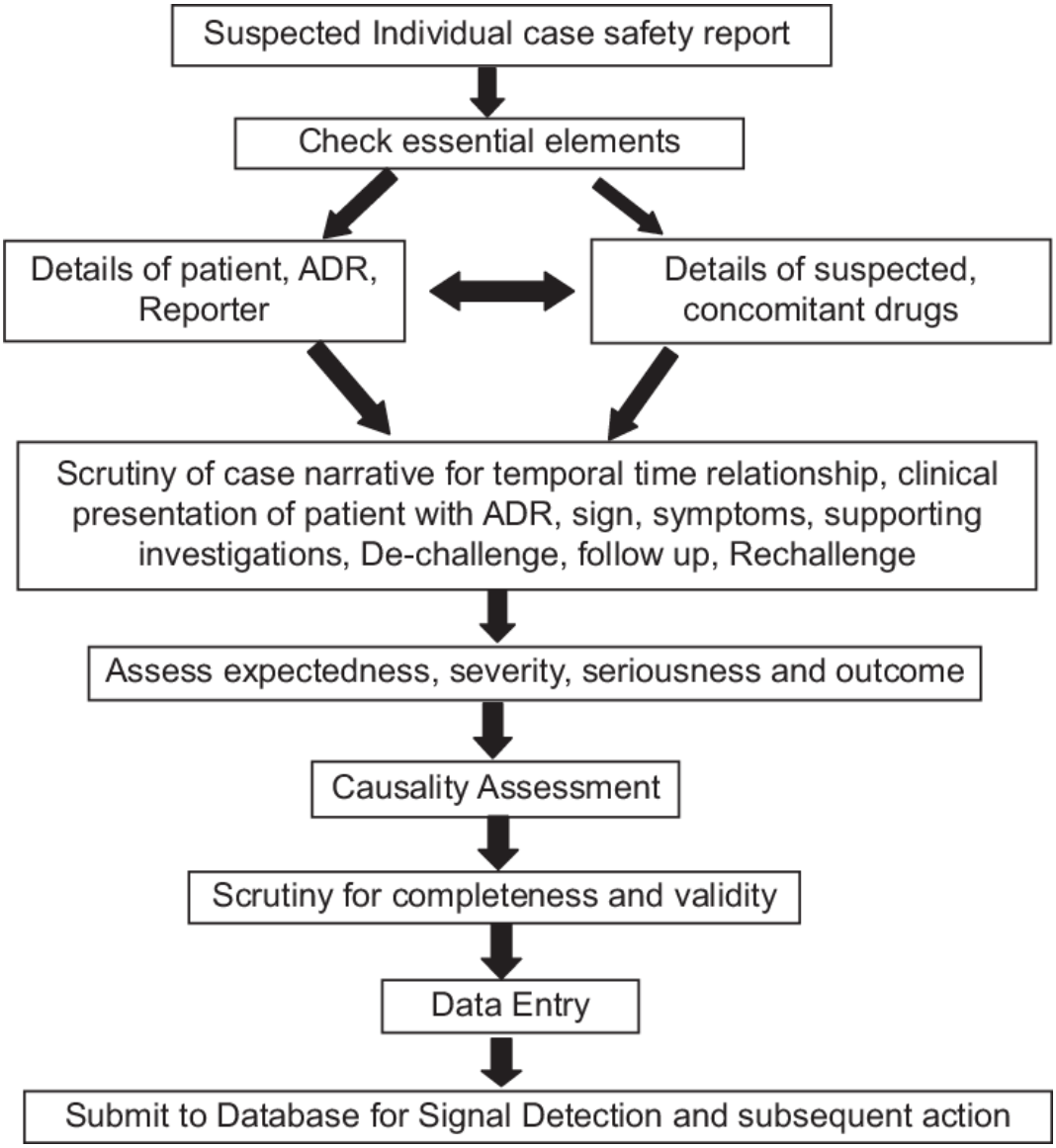

Figure 1 A functioning pharmacovigilance system. Individual case safety reports (ICSRs) are collected by healthcare professionals and patients worldwide, then coded and analyzed to detect drug safety signals. (CC BY-SA 4.0)

Data Sources for Pharmacovigilance

Spontaneous Reporting Systems (SRS)

The cornerstone of traditional PV is spontaneous reporting. National agencies and healthcare providers submit suspected ADR reports to centralized databases. Major examples include the U.S. FDA Adverse Event Reporting System (FAERS) and the WHO Programme for International Drug Monitoring's VigiBase. These systems capture suspected ADRs for a wide range of drugs, biologics, and vaccines. For instance, FAERS includes both ADRs and medication errors reported in the U.S. The WHO VigiBase aggregates reports from over 160 countries, accounting for roughly 99% of the world's population. As of mid-2023, VigiBase contained over 35 million reports (World Health Organization., 2023).

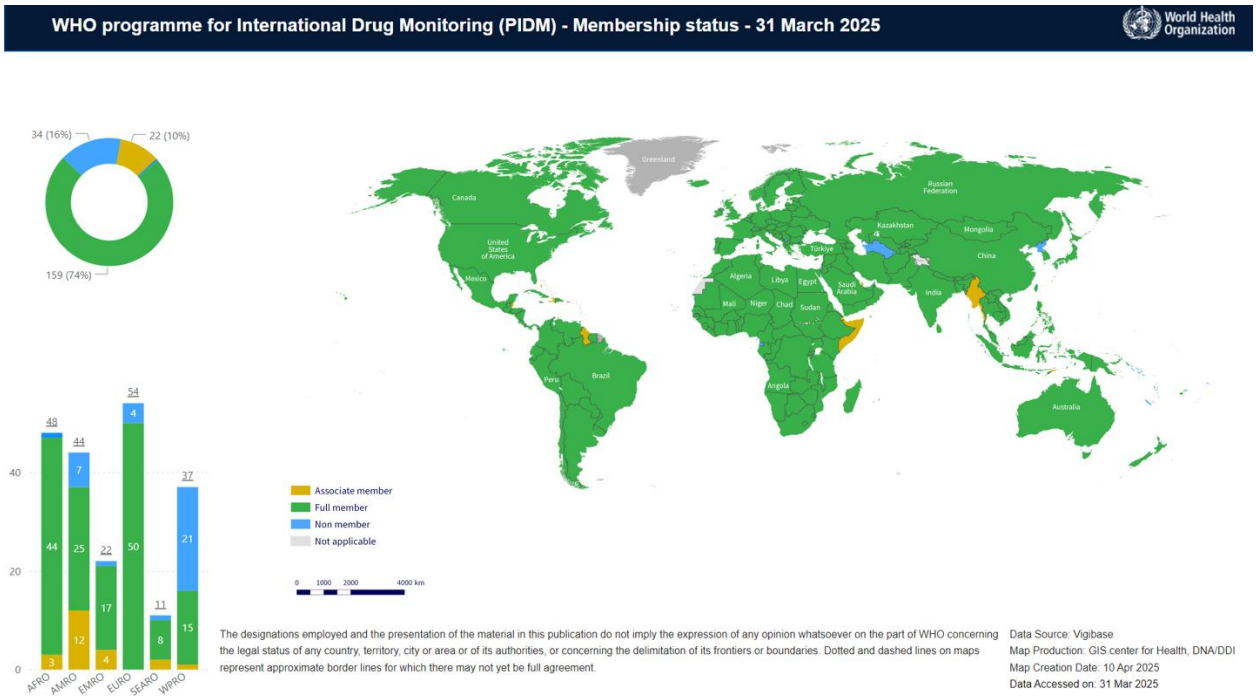

Figure 2 Membership status of countries in the WHO Programme for International Drug Monitoring (PIDM) as of 31 March 2025. While the number of reports (over 35 million) reflects 2023 data, this map shows updated membership coverage contributing to VigiBase.

Spontaneous reporting has strengths and weaknesses. It can identify rare or unexpected reactions that were not seen in clinical trials, but it suffers from under-reporting and reporting biases. Estimates suggest only a small fraction of all ADRs are reported this way. Nonetheless, SRS data can be mined with statistical methods. Classic disproportionality measures such as reporting odds ratios (ROR) and proportional reporting ratios (PRR) are used to flag drug-ADR pairs reported more often than expected. For example, one study notes that national SRS are "particularly useful for detecting unexpected and unknown adverse events".

Electronic Health Records and Claims

Beyond SRS, electronic medical records offer complementary real-world data (RWD). EHR systems record patient diagnoses, lab results, medications, and outcomes. Claims databases track prescriptions and hospitalizations over time. Mining EHR and claims data allows analysis of ADRs in routine care settings, capturing events that may never be reported spontaneously. For example, insurance claims data (such as the FDA's Sentinel System) support active surveillance of drug safety by identifying cohorts of exposed patients. One recent review found that combining EHR data with SRS and claims often improves detection compared to any single source.

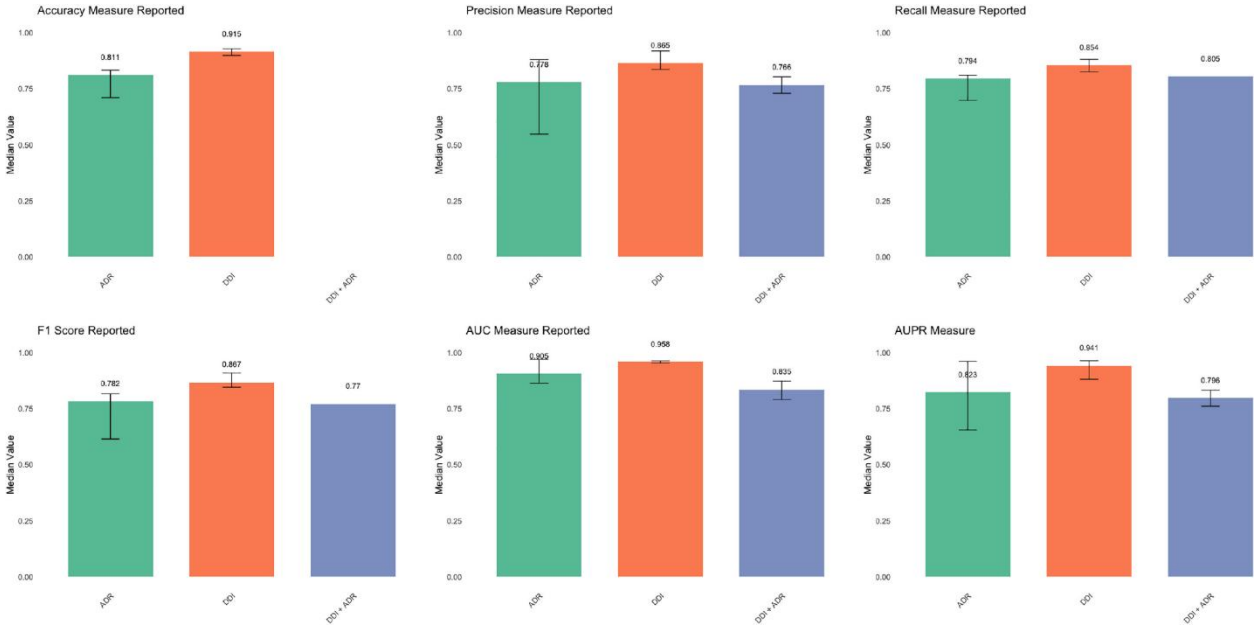

Figure 3 Comparison of machine learning model performance in detecting adverse drug reactions (ADRs), drug–drug interactions (DDIs), and combined DDI+ADR datasets. Metrics include accuracy, precision, recall, F1 score, AUC, and AUPR. Results show improved performance when combining EHR-based ADR and DDI data. Adapted from Kim, J. V., et al. (2024). Frontiers in Drug Safety and Regulation.

However, using RWD comes with challenges. EHR data are often unstructured and fragmented. Adverse events may not be explicitly linked to drugs, requiring natural language processing or complex phenotype definitions. Nonetheless, recent research (e.g. Kim *et al.*, 2024) has developed methods to extract ADR signals from EHRs by leveraging diagnosis codes, laboratory values, and clinical notes. Such approaches can validate known ADRs or suggest new ones.

Novel Data Sources

Besides traditional and EHR data, new sources of patient-generated health data are emerging. These include social media posts, patient forums, and even search engine queries. For example, analyses of Twitter or health forums can pick up patient-reported side effects earlier or highlight subjective experiences of ADRs. Mobile health apps and wearable devices also represent potential data streams. These sources are still experimental but point to a broader data ecosystem.

Publicly Available Datasets

For research, some PV datasets are publicly accessible. The FDA FAERS data can be downloaded through openFDA APIs, enabling custom analysis of U.S. ADR reports. Similarly, WHO's VigiAccess portal provides aggregate statistics, and UMC has released selected data and tools. Pharmacovigilance researchers have also created reference datasets: for example, Lee *et al.* (2022) developed a "reference standard for ADR signal assessment" by integrating multiple ADR reference sets.

Table 1 Data sources in pharmacovigilance and examples of use.

Data Source	Content	Use Case	Example Reference
Spontaneous Reports	ICSRs (VigiBase, FAERS)	Signal detection via disproportionality	WHO (2023); Park et al. (2022)
Electronic Health Records	Patient diagnoses, labs, notes	Cohort studies, signal validation	Kim et al. (2024)
Health Insurance Claims	Billing codes, prescriptions	Active surveillance, epidemiology	Davis et al. (2023)
Social Media & Web Data	Posts, search queries, patient forums	Early signal detection, patient insights	Lee et al. (2021)

Data-Driven Methods

Statistical Signal Detection

Traditional PV relies on statistical screening of large ADR datasets. Disproportionality methods remain popular: for example, calculating the reporting odds ratio (ROR) for each drug-event pair to find outliers. Park *et al.* (2022) review notes that in FAERS and similar SRS, "disproportionality analysis through FAERS" is a common technique. Other methods include Bayesian measures (e.g. the MGPS or BCPNN algorithms) and regression-based models. These methods generate hypotheses about drug safety that must be reviewed by experts.

Machine Learning and AI

More recently, machine learning (ML) and artificial intelligence (AI) are applied to PV. ML can handle high-dimensional data and find complex patterns in reports and text. Desai (2024) emphasizes that the growth in volume and variety of ADR reports has prompted "a lot of excitement and enthusiasm to adopt AI technology to automate PV". For example, natural language processing (NLP) can extract drug-ADR relations from narrative reports or scientific literature. Classification algorithms (random forests, neural networks) can predict the probability that a given adverse event report is a true ADR.

AI has been used at various PV steps. Some tools triage ICSRs by estimating missing data elements or coding inaccuracies. Others build predictive models that use patient demographics and drug properties to forecast likely ADRs. The systematic review by Park *et al.* (2022) found many recent studies applying ML to ADR data, using methods like support vector machines and deep learning on molecular or phenotypic features. However, these approaches depend heavily on high-quality labeled data, which remains a limiting factor.

Reference Standards and Validation

A key issue in data-driven PV is validation. Many signal-detection methods produce false positives/negatives. One strategy is to use reference sets of known positive and negative drug-ADR pairs. For instance, Lee et al. (2022) created a "reference standard for ADR signal assessment" by merging existing resources (SIDER, OMOP, EU-ADR) and mapping them to EHR data. Such reference sets allow systematic evaluation of new algorithms. The same review by Park et al. (2022) highlights that "validating massively detected ADR signals is a major challenge" and requires curated knowledge bases.

Example Analysis (FAERS Data)

As an illustrative example, consider analyzing publicly available FAERS data for a specific drug class (e.g. antibiotics). Using openFDA, one can query FAERS for all reports where drugs in that class are the primary suspect. A simple disproportionality analysis would compare the frequency of each reported event to its background frequency. For example, (hypothetically) if out of 10,000 antibiotic-associated reports, 500 involve rashes, while in all FAERS 100,000 reports rashes appear 1,000 times, the ROR for rash would be (500/9,500) / (1,000/99,000) ≈ 5.47, suggesting a signal. In practice, one must control for confounders and reporting biases.

Researchers have done similar studies. For instance, openFDA data have been mined for signals of liver toxicity or cardiovascular events associated with drug classes. Open-source tools like *OpenVigil* facilitate such analyses (an example of practical PV tool). The FAERS Public Dashboard by FDA also shows charts of top reported reactions by drug class (U.S. Food and Drug Administration., 2023). These resources exemplify how publicly available PV data can be used for quantitative safety monitoring.

Regional and Demographic Patterns

ADR patterns can vary by region, healthcare system, and patient population. Wakao *et al.* (2019) performed a global comparison of ICSRs and found striking differences. For example, Japan's PV system submitted many reports from physicians and pharmacists, whereas U.S. reports had a higher share from consumers. They also found that elderly patients (ages 70–89) were more represented in Japanese reports, and some ADRs (interstitial lung disease, abnormal liver function) were disproportionately reported in Japan. Such pharmaco-ethnic signals are important for tailoring drug safety efforts in different countries.

Globally, ADR rates depend on drug use patterns and monitoring practices. When pooling VigiBase data, one often sees that antibiotics, analgesics, and cardiovascular drugs account for a large share of ADR reports. For instance, one global analysis noted that vaccines and antibiotics are among the most reported categories. Surveillance systems also track drugs of recent concern: e.g., rollout of new biologics is accompanied by intense PV scrutiny

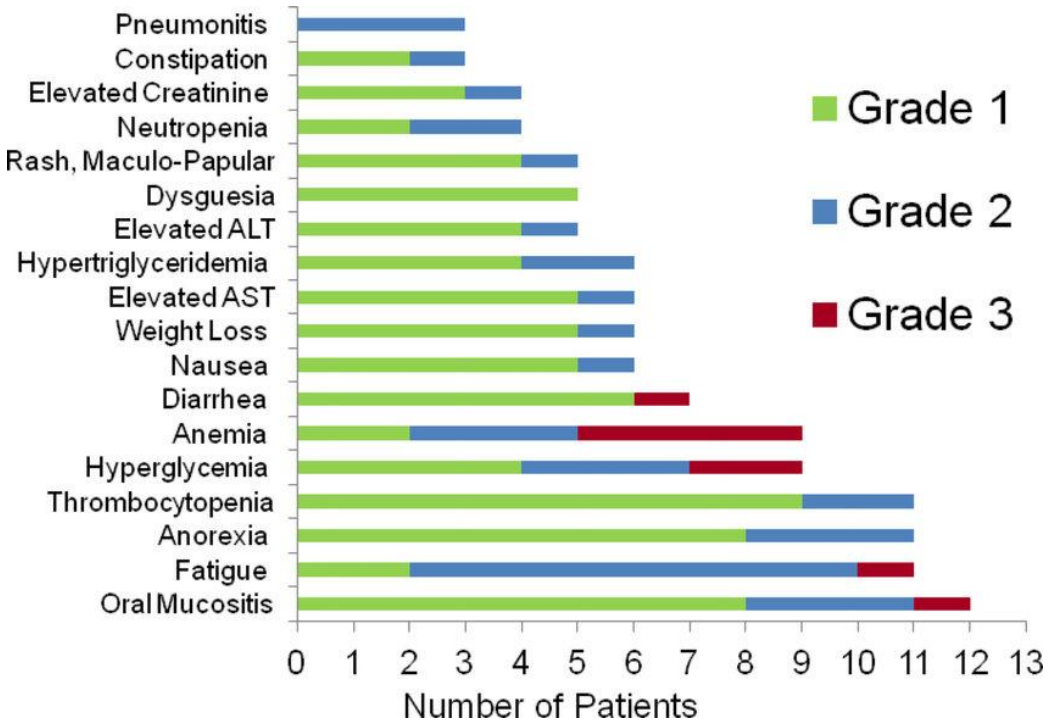

Figure 4 Severity distribution of adverse drug reactions (ADRs) by type and patient count. The chart categorizes 19 ADRs based on severity: Grade 1 (mild, green), Grade 2 (moderate, blue), and Grade 3 (severe, red). Oral mucositis, fatigue, anemia, and hyperglycemia showed the highest number of severe (Grade 3) cases.

.

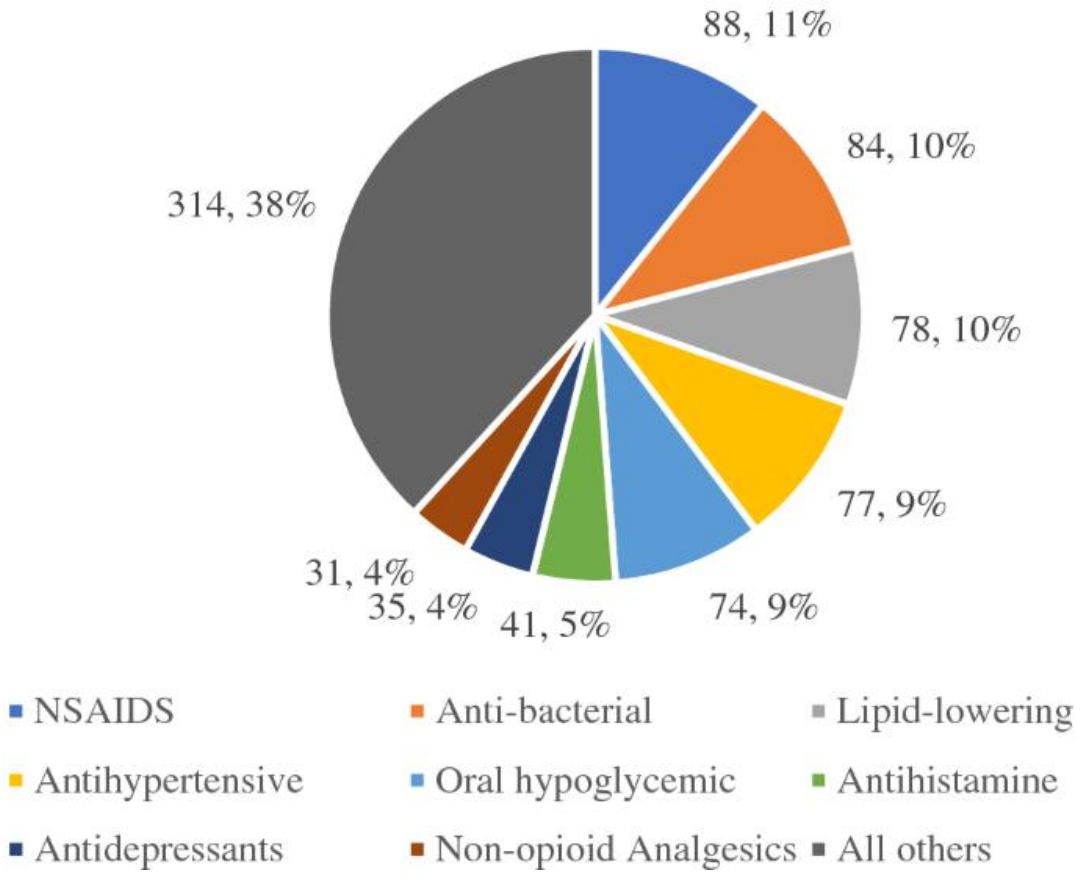

Figure 5 Distribution of adverse drug reaction (ADR) reports by drug category.

The chart shows that "All others" account for 38% of ADRs, followed by NSAIDs (11%), anti-bacterial agents (10%), lipid-lowering drugs (10%), antihypertensives (9%), oral hypoglycemics (9%), antihistamines (5%), antidepressants (4%), and non-opioid analgesics (4%). (Adapted from Braun, T., et al. (2020). Drugs-Real World Outcomes, 7, 167–176)

Outcomes and Preventive Measures

The ultimate goal of pharmacovigilance is to improve patient safety. When signals emerge, regulators and clinicians can take action: updating drug labels, issuing warnings, or restricting use. Data-driven PV also enables feedback loops. For example, if an EHR study finds that a specific drug causes severe reactions in a demographic group, guidelines can be changed. In India, an educational intervention dramatically improved ADR reporting rates in a hospital, illustrating how awareness affects data quality (Lavertu et al., 2021).

Innovations in data analytics have led to proactive risk management. Large-scale mining can detect shifts in ADR patterns early. Some countries now use mobile apps to encourage patient reporting, expanding the PV base. The collaboration of international PV programs, combined with advanced data science, holds promise to detect even subtle drug risks before they harm many patients.

Challenges and Future Directions

Despite progress, pharmacovigilance faces ongoing challenges. Underreporting remains a major issue: many ADRs are never documented, especially mild ones. Data may be incomplete or inconsistent across sources. Integrating heterogeneous data (SRS, EHR, registries) requires standardization efforts like the OMOP Common Data Model help (Shin et al., 2021). Ensuring data privacy is also crucial.

AI and machine learning offer solutions but come with caveats. Desai (2024) warns that full automation of PV is a “double-edged sword”: models must be validated and transparent, and human oversight is needed. Ethical and practical issues (e.g. false alarms or missing signals) must be addressed. Ongoing research in “explainable AI” and collaborative platforms (like open-source PV dashboards) will be important.

Looking forward, the PV community is increasingly adopting real-world evidence. Regulatory agencies now accept RWD (from EHRs, wearables, etc.) as supplementary safety evidence (Lavertu et al., 2021). The shift is towards continuous monitoring: rather than passive surveillance, systems will use streams of data to update safety profiles in near real-time. This is a “new era” of pharmacovigilance.

Conclusion

Data-driven pharmacovigilance leverages large-scale healthcare data and computational methods to detect and understand adverse drug reactions in clinical practice. By mining sources from international ADR databases to electronic health records, researchers and regulators can uncover patterns that inform safer use of medications. Recent studies illustrate how disparities in reporting (e.g. across countries) and advanced analytics (signal scores, machine learning) contribute to drug safety knowledge. As the volume of health data continues to grow, sophisticated PV algorithms will be essential to separate meaningful safety signals from noise. Ultimately, combining quantitative analytics with expert review promises more proactive drug safety monitoring and better patient outcomes worldwide.

References

1. Desai, M. K. (2024). Artificial intelligence in pharmacovigilance - Opportunities and challenges. Perspectives in Clinical Research, 15(3), 116–121. https://doi.org/10.4103/picr.picr_290_23
2. Kim, J. V., Davis, S. E., Matheny, M. E., & Smith, J. C. (2024). Integrating electronic health records with other data sources for postmarket drug safety signal identification: A review. Frontiers in Drug Safety and Regulation, 4, Article 1428831.
3. Lavertu, A., Bouchard, B., Flamand, V., Paquet, F., Labrosse, É., Beauchesne, L., & L’Heureux, S. (2021). A new era in pharmacovigilance: Toward real-world data and digital monitoring of medicine safety. *Clinical Pharmacology & Therapeutics*, 110(4), 1194–1204. https://doi.org/10.1002/cpt.2545
4. Lee, S., Lee, J. H., Kim, G. J., Kim, J.-Y., Shin, H., Ko, I., Choe, S., & Kim, J. H. (2022). A data-driven reference standard for adverse drug reaction (RS-ADR) signal assessment: Development and validation. *Journal of Medical Internet Research*, 24(10), e35464. https://doi.org/10.2196/35464
5. Park, S., Lee, H., & Park, H. (2022). Analyzing adverse drug reaction using statistical and machine learning methods: A systematic review. *Medicine*, 101(36), e30098. https://doi.org/10.1097/MD.0000000000030098

6. Shin, J. M., Chung, W. I., Kwon, J., & Jee, C. (2021). Data-driven pharmacovigilance: A text mining and pharmacovigilance approach for predicting adverse drug events. *Applied Sciences*, 11(5), 2249. https://doi.org/10.3390/app11052249
7. Hamza Khalifa Ibrahim (2025). Global Trends in Cardiovascular Mortality and Risk Factors: Insights from WHO and Global Burden of Disease Data. (2025). Libyan Open University Journal of Medical Sciences and Sustainability, 1(1), 28-36
8. Wakao, R., Taavola, H., Sandberg, L., Iwasa, E., Soejima, S., Chandler, R., & Norén, G. N. (2019). Data-driven identification of adverse event reporting patterns for Japan in VigiBase, the WHO global database of individual case safety reports. *Drug Safety*, 42(12), 1487–1498. https://doi.org/10.1007/s40264-019-00861-y
9. World Health Organization. (2023, July). *The WHO Programme for International Drug Monitoring*. https://www.who.int/teams/regulation-prequalification/regulation-and-safety/pharmacovigilance/networks/pidm
10. *U.S. Food and Drug Administration*. (2023). *FDA Adverse Event Reporting System (FAERS) Public Dashboard*. https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard
11. Braun, T., Griese-Mammen, N., Schneider, A., & Schulz, M. (2020). Drug-related problems in older adults: A classification and literature review. Drugs-Real World Outcomes, 7(3), 167–176.
12. Nusaiba Abduladim Elmahjub (2025). Global Trends in Prescription Drug Utilization and Resistance Patterns: A 15-Year Comparative Study of Therapeutic Practices. (2025). Scientific Journal for Publishing in Health Research and Technology, 1(1), 26-34.
13. Aejeeliyah Yousuf (2025). Assessment of Antibiotic Resistance Trends Across Regions Through a 10 Year Retrospective Data Analysis. (2025). Scientific Journal for Publishing in Health Research and Technology, 1(1), 17-25.
14. Nusaiba Abduladim Elmahjub (2025). Deep Learning for Early Detection and Classification of Diabetic Retinopathy in Clinical Practice. (2025). Libyan Open University Journal of Medical Sciences and Sustainability, 1(1), 11-19.