Predicting Near-Term Adverse Drug Events in Polypharmacy: An Opioid-Focused Claims-Based Study

Benet Fité Abril, Yunxi Li, Ankit Pal

1 Abstract

This study presents a scalable pipeline for transforming administrative claims data into a format suitable for identifying polypharmacy-related adverse drug events (ADEs) and evaluating their short-term predictability. Using Inovalon's enterprise claims database, we constructed longitudinal "polypharmacy spells" and defined Time 0 as the moment a medication was added to a regimen of three or more concurrent drugs. For a focused use case, we analyzed opioid-involved spells due to their high clinical relevance [1]. The final cohort contained more than 800,000 spells from $\tilde{2}$00,000 patients.

Despite substantial class imbalance (1.3% ADE rate), machine learning models captured meaningful clinical signal. XGBoost modestly outperformed logistic regression (AUC-PR 0.0518 vs. 0.0441) and achieved a 15-fold enrichment of ADEs among the highest-risk predictions, though performance remained below clinical utility thresholds. Explainability analyses showed the models primarily detected markers of patient complexity rather than opioid-specific interactions. However, statistical tests identified several drug combinations with significantly elevated ADE incidence, including known opioid interaction partners.

This work establishes a reproducible pipeline for large-scale polypharmacy analysis and demonstrates the feasibility of near-term ADE risk prediction from claims data, while highlighting the need for expanded cohorts and richer clinical context to achieve deployable models.

DISCLAIMER: ChatGPT [2] was used throughout this project as a coding assistant and literature researcher. In this report, no Generative AI is replacing our own intellectual effort, ChatGPT has only been used for rewriting, polishing and condensing information.

2 Introduction

Polypharmacy, the concurrent use of multiple medications, is increasingly common in modern medicine and has been associated with elevated risk of adverse drug events (ADEs) [3][4]. Although prior work has characterized the burden of polypharmacy and documented ADE frequency, far less is known about how specific drug combinations, patient characteristics, and regimen changes contribute to short-term ADE risk, particularly when medication regimens are modified. Understanding these interactions could enable clinicians to mitigate harm by identifying high-risk prescribing patterns at the point of medication change, supporting integration into electronic health record (EHR) decision support systems.

The objective of this study is (1) to develop a reusable data processing pipeline, (2) explore clinical and pharmacological factors associated with near-term ADEs following drug combination changes, and (3) evaluate whether machine learning models can predict these events from large-scale claims data. To ground the investigation in a clinically meaningful setting, we selected polypharmacy patients taking opioids, due to their high prevalence in the US and high rate of ADEs [1].

3 Methods

Eligibility Criteria: At the beginning of the project, our aim was too broad, as we intended to focus on all elderly polypharmacy patients. This yielded a very sparse cohort, so based on the preliminary results we decided to change the scope to members with $\geq$180 days of continuous pre-index enrollment (to balance comorbidity capture and cohort size), with $\geq$3 concurrent drugs including **at least one opioid** and without previous history of ADEs to avoid confounding. We removed the age restriction due to fewer than expected elderly patients in the cohort.

Time Zero Definition: Due to the trends observed (see Results), which showed that many ADEs were occurring concurrently with drug additions to a patient's prescription, we defined Time 0 as each time a patient added a drug to their combination, as long as the eligibility criteria were still met. The prediction target was defined as the occurrence of a new ADE within 30 days of this event.

Data Source and Extraction: We used 1M and 5M patient samples from the Inovalon enterprise claims ecosystem, comprising prescription fills, inpatient/outpatient claims, ICD-10 diagnoses, demographics, and enrollment periods. ADE ICD-10 codes were curated from previous research [5] identifying the codes more associated with ADEs, prioritizing classes A1, A2, B1, B2, and C. The selected list can be found here. For the selection of opioid drugs, National Drug Codes (NDC11) were downloaded from OpenFDA [6], filtered to opioid ingredients and normalized through ndclib-based conversions [7]. The selected list can be found here.

The final pipeline extracts raw data from SQL servers filtering members based on the eligibility criteria, enables selective drug and ADE identification, and stores the data in parquet files in the O2 cluster. Due to traffic in the SQL database, we decided to offload as much processing as possible to O2 and python, avoiding the SQL bottleneck.

Before building the polypharmacy spells, drugs were clustered at the ATC3 class level to avoid artificially distinct representations of dose forms. NDC11 codes were first mapped to RxNorm using NDC-RxNorm crosswalk tool, then mapped to ATC3 using the sagerx tool [8, 9]. Comorbidities were clustered at the highest ICD-10 level (eg. A00).

Spell Construction: After the raw data is downloaded into the cluster, a series of pandas-based python scripts build the polypharmacy spells using concurrent drug intervals, merging consecutive same-drug intervals, enforcing the eligibility criteria and allowing a grace period once counts drop below the minimum concurrent drugs, to account for prescription renewal delays. Since we were not working with time-to-event endpoints, we decided to set a minimum spell duration to remove short non-informative spells. In Figure 1 the polypharmacy spell generation process is represented. Afterwards, the spells are split by drug additions to get our final time 0 aligned cohort and data cleaning is performed to maximize quality and avoid data leakage.

The detailed instructions to run the pipeline can be found here, with detailed information on the output tables. For the opioid use case, we chose a grace period of 15 days and a minimum spell duration of 15 days.

Computational challenges: Due to the large size of the dataset, the code had to be optimized for performance, chunking, parallelizing and avoiding loading large tables into memory where possible. It is now possible to run the whole pipeline in around 10 hours on the 5M database, depending on SQL traffic. To scale to the full Inovalon dataset, the processing strategy would need to be refined.

Another challenge was ensuring the final tables were actually representing what was intended. Many data quality issues and bugs were identified in downstream EDAs and modeling results and subsequently corrected, yielding a pipeline robust to data leakage.

Modeling Approaches: The objectives of the modeling were twofold: (1) to identify opioid drug combinations that are more strongly associated with ADE than baseline, and (2) to evaluate whether the incidence of ADE can be predicted within 30 days of a medication change. Our data set simulates a clinically actionable setting, where a predictive model could be triggered at the moment a new drug is added to a regimen, aligned with our Time 0 definition.

To identify drug combinations with a significantly higher risk of ADE than average, we used **pairwise adjusted BH chi-square proportion tests**. It does not correct for comorbidity confounders, but can provide interesting directions for future in-depth causal studies on the full Inovalon dataset.

For prediction, both classical ML and DL methods were explored. The predictors used were comorbidities, drugs on prescription, age, gender and race. For ML, multi-hot encoding was used for drugs and comorbidities, and one-hot encoding was used for demographics. The approaches explored in this study are:

- **Baseline logistic regression**: interpretable and baseline predictive benchmark, allows for identification of high odds-ratio factors with confounder adjustment. To account for imbalance, we used l2 regularization, the saga solver and a class-weighted loss function.
- **XGBoost**: gradient-boosted trees with strong performance in sparse tabular datasets, which allows for the use of SHAP to identify the weight of specific variables on decision trees. To account for imbalance, we optimized the parameter scale_pos_weight to $0.2 * ADEprevalence$.
- **Multilayer Perceptron + MedTok**: a basic but powerful DL model that allowed us to try MedTok [10], a library that provides pretrained embeddings for the ATC classes and ICD comorbidities based on a knowledge graph. Multiple architectures for the embeddings (flatten, mean, attention) and various hidden layer sizes were explored.

For the evaluation, we opted for:

- **AUC-PR** - The primary performance metric due to the very low prevalence of events, as it focuses on how well the model identifies the minority class. Interpreting AUC-PR relative to the ADE prevalence allows us to quantify how much better than random guessing the model is performing.
- **AUC-ROC** - Included as a secondary metric, although it is not optimal for highly imbalanced datasets.
- **Precision@k**: to see the ability of the models to rank positive cases higher than negative ones.
- **Recall@x precision**: to put the real-world impact of the predictive algorithms in easily understandable terms: *if I want a precision of x%, what will my recall be?*

Generalizability of method: The polypharmacy analysis pipeline developed here is adaptable to a wide range of research objectives. It can support broad investigations such as the opioid-focused use case explored in this study, or highly targeted analyses of specific drug combinations or ADE types. Its modular structure facilitates extension, and further engineering work could enable deployment across Inovalon's 200M claims database.

4 Results

Cohort Overview: The final cohort represented in the CONSORT diagram in Figure 2 included 814,412 opioid-involved polypharmacy spells from 206,664 patients, of which 10,898 (1.3%) suffered an ADE within

30 days of a drug change. Demographic and medication-pattern distributions, including age, gender, race and comorbidity counts, were consistent with expectations for a high-utilization population (Table 1), indicating no systematic biases in cohort construction.

Analysis of ADE patterns revealed a pronounced temporal signal: after aligning spells by Time 0, there was a significant data shift: $\tilde{8}0\%$ of ADEs occurred within 30 days of the regimen change, and most were opioid- or dosage-related (Figure 3), vs only $\tilde{3}5\%$ in the non-splitted data. This concentration of clinically coherent events immediately following medication adjustments supports the relevance of our outcome definition and motivates the use of predictive models targeting near-term ADE risk.

In the annex you can find lists with the drug combinations associated with a higher ADE prevalence, with the relevant chi-sq p-value compared against the rest of the cohort (Table 2, 3).

Model Performance and interpretability: With a baseline risk for a 30-day ADE of 1.3%, this was a challenging prediction task. Logistic regression modestly exceeded this baseline, achieving an AUC-PR of 0.0441 and concentrating 16.0% ADE prevalence in the top 0.1% of alerts. XGBoost delivered the strongest performance, increasing AUC-PR to 0.0518 and achieving 20.3% precision in the top 0.1% of predictions, a 15x enrichment over baseline (Figure 4). The neural MLP-MedTok approach underperformed the tree-based model, and would require more architectural tuning and probably more data to ensure enough signal.

Feature attribution analyses using odds ratios (logistic regression) and SHAP values (XGBoost) revealed distinct clinical and pharmacological drivers of model predictions (Figures 4, 5), which are explored in greater detail in the Discussion.

5 Discussion

First of all, the analysis of the opioid cohort and the interpretability metrics of the models allowed us to validate the robustness and correctness of the pipeline. Proportion-based test results were consistent with reported drug-drug interactions [11]: antiemetics (A04A) have been reported to cause enhanced opioid exposure and sedation, and antipsychotics (N05A) have been reported by the FDA to have risk of synergistic respiratory depression and risk of overdose [12]. Most other high p-value pairs, for example chemotherapy agents (L01X/L01E/L01A), likely reflect complex pharmacokinetic interactions, contextual acuity (ICU stays) and general patient frailty, rather than specific interaction with opioids.

Although the predictive performance of our models remains modest by clinical standards, the observed enrichment at the high end of the risk spectrum indicates that adverse drug events are not randomly distributed in this population and that the database contains some genuine pharmacologic and clinical signal, even though the performance is far from a clinically deployable model.

Some limitations that could explain the low performance come from claims data itself and from the study design. Regarding claims data, the main issue is that it lacks dosage timing, so so we made a general assumption of one dose per day. This could be improved in studies with a narrower set of drugs by curating a table of dosage / drug. Also, as it is common in pharmacy fills, you never know if the patient actually took the prescription or not. The dataset also does not capture non-prescription drugs such as NSAIDs, which could interact with other drugs and confound results.

Regarding the study design, there are two main limitations: clustering and cohort size. For drugs, we are clustering at the ATC3 level, as we are covering a broad range of drugs and allowed to identify general associations. Ideally a more narrow clinical question should cluster at more granular levels, to avoid loss of information. For comorbidities, we are clustering ICD codes to the highest level of the tree. More complex

approaches such as PheCodes [13] or other clinically meaningful comorbidity scores should be explored in future studies. With respect to the cohort size, scaling to the 200M dataset would allow deep learning approaches, leveraging MedTok to encode the drugs and comorbidities in a meaningful way as we showed in our proof of concept. The pre-index window could also be extended, even if at risk of adding noise to the models.

The interpretability plots provide a final layer of insight towards the interpretation of the results. Both logistic regression and XGBoost mostly focus on drugs and comorbidities that act as proxies of general patient complexity and general ADE predictors, rather than specific opioid related interactions or factors. This suggests that predicting ADEs within an opioid-focused polypharmacy cohort may not have been the optimal design choice, and suggests two possible directions for improving prediction are:

- Go back to our initial hypothesis and try to generally capture the indicators of patient complexity and frailty on a general dataset of polypharmacy, allowing a big deep learning model to cluster and separate the patients based on their comorbidities and drug prescriptions. This approach however risks predicting that *sicker patients get more ADEs*, which is not that clinically useful.
- Narrow down the definition of what opioid patient profile we are focusing on (chronic pain vs cancer patients vs addictive disorders, etc) and develop more homogeneous class-specific models, to see if the predictions improve on the stratified groups.

Learnings and personal reflection: This project allowed us to battle with the complexities of claims data, exposing the temporal power of this kind of data and its shortcomings. It also taught us that having a clear clinical question is very important to be able to produce meaningful results, as a common struggle during the semester was the lack of specificity of the clinical question and the constant refining of it.

Next steps: Apart from the two possible study design variations described in the discussion, this project's insights point towards an interesting series of next steps:

- **Scale up**: adapt the code to work on the 200M database, leveraging SQL and the O2 resources efficiently.
- **Pharmacoepidemiology drug-drug interactions**: more interesting than predicting that sicker patients get more ADEs, the pipeline could be reused to do sensitivity analysis on drug combinations, using specific ADE occurrence as an endpoint. One example of question that could be answered with small tweaks to the pipeline is: Among adults with chronic non-cancer pain already receiving long-term opioid therapy, what is the effect of initiating gabapentin vs initiating duloxetine as an adjunct analgesic on the risk of serious opioid-related adverse drug events (overdose, respiratory depression, ED visit, hospitalization)? [14, 15]
- **Merging with EHR**: A model predicting ADEs would be deployed in an EHR, so maybe leveraging the increased information density of the EHR would yield better results.
- **Automation**: Claims data is very difficult to work with, both due to the format of the data itself and the size of it. User-friendly AI wrapping of the pipeline could allow professionals with less computational knowledge to run preliminary cohort size assessment and exploration, and even run simple end-to-end analysis.

References

[1] E. Y. Liu, K. L. McCall, and B. J. Piper, "Variation in adverse drug events of opioids in the United States," *Frontiers in Pharmacology*, vol. 14, p. 1163976, Mar. 2023.

[2] OpenAI, "ChatGPT." https://chatgpt.com/. Accessed: Nov. 29, 2025.

[3] J. Komagamine, "Prevalence of urgent hospitalizations caused by adverse drug reactions: a cross-sectional study," *Scientific Reports*, vol. 14, p. 6058, Mar. 2024. Publisher: Nature Publishing Group.

[4] Z. Wang, T. Liu, Q. Su, H. Luo, L. Lou, L. Zhao, X. Kang, Y. Pan, and Y. Nie, "Prevalence of Polypharmacy in Elderly Population Worldwide: A Systematic Review and Meta-Analysis," *Pharmacoepidemiology and Drug Safety*, vol. 33, p. e5880, Aug. 2024.

[5] C. M. Hohl, A. Karpov, L. Reddekopp, and J. Stausberg, "ICD-10 codes used to identify adverse drug events in administrative data: a systematic review," *Journal of the American Medical Informatics Association : JAMIA*, vol. 21, pp. 547–557, May 2014.

[6] U.S. Food and Drug Administration, "openfda Drug NDC API." https://open.fda.gov/apis/drug/ndc/download/. Accessed: Nov. 29, 2025.

[7] E. Cosma, "eddie-cosma/ndclib." https://github.com/eddie-cosma/ndclib, Mar. 2025. MIT License. Accessed: Nov. 29, 2025.

[8] National Library of Medicine, "RxNorm Technical Documentation: NDC to RxNorm Crosswalk." https://www.nlm.nih.gov/research/umls/rxnorm/docs/techdoc.html#s12_5. Section 12.5: NDC to RxNorm Code Crosswalk.

[9] CoderXio, "SageRx: ATC Codes to RxNorm Products Mapping." https://coderxio.github.io/sagerx/#!/model/model.sagerx.atc_codes_to_rxnorm_products. SageRx Data Model Documentation.

[10] X. Su, S. Messica, Y. Huang, R. Johnson, L. Fesser, S. Gao, F. Sahneh, and M. Zitnik, "Multimodal Medical Code Tokenizer," June 2025. arXiv:2502.04397 [cs].

[11] X.-q. Feng, L.-l. Zhu, and Q. Zhou, "Opioid analgesics-related pharmacokinetic drug interactions: from the perspectives of evidence based on randomized controlled trials and clinical risk management," *Journal of Pain Research*, vol. 10, pp. 1225–1239, May 2017. Publisher: Dove Press.

[12] A. G. Szmulewicz, B. T. Bateman, R. Levin, and K. F. Huybrechts, "Risk of Overdose Associated With Co-prescription of Antipsychotics and Opioids: A Population-Based Cohort Study," *Schizophrenia Bulletin*, vol. 48, pp. 405–413, Sept. 2021.

[13] P. Wu, A. Gifford, X. Meng, X. Li, H. Campbell, T. Varley, J. Zhao, R. Carroll, L. Bastarache, J. C. Denny, E. Theodoratou, and W.-Q. Wei, "Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation," *JMIR Medical Informatics*, vol. 7, p. e14325, Nov. 2019. Company: JMIR Medical Informatics Distributor: JMIR Medical Informatics Institution: JMIR Medical Informatics Label: JMIR Medical Informatics Publisher: JMIR Publications Inc., Toronto, Canada.

[14] J. Hahn, Y. Jo, S. H. Yoo, J. Shin, Y. M. Yu, and Y.-M. Ah, "Risk of major adverse events associated with gabapentinoid and opioid combination therapy: A systematic review and meta-analysis," *Frontiers in Pharmacology*, vol. 13, Oct. 2022. Publisher: Frontiers.

[15] T. Gomes, D. N. Juurlink, T. Antoniou, M. M. Mamdani, J. M. Paterson, and W. van den Brink, "Gabapentin, opioids, and the risk of opioid-related death: A population-based nested case–control study," *PLoS Medicine*, vol. 14, p. e1002396, Oct. 2017.

6 Acknowledgements

We would like to thank Inovalon for allowing us to access their datasets. All work was performed as part of Harvard's BMIF204 course, taught by Dr. Sebastian Schneeweiss.

7 Reproducibility

Code is available at `https://github.com/beni-1414/bmif204_claims_project`.

A Figures and tables

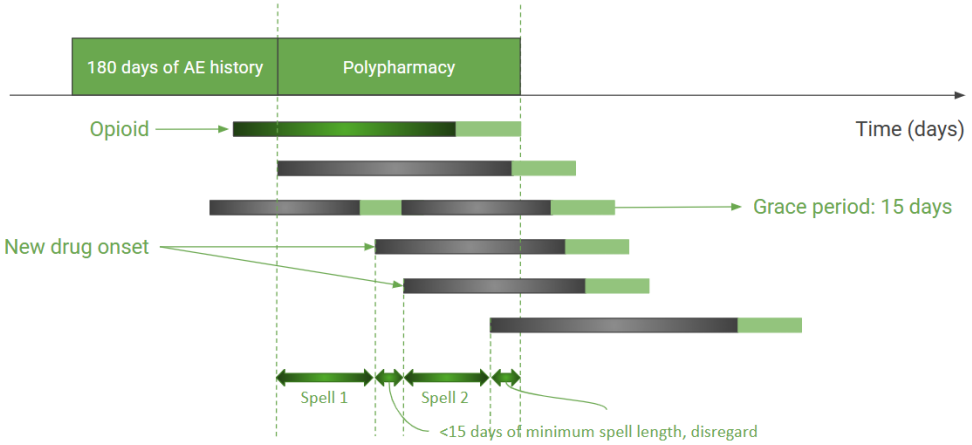


Figure 1: Representation of the spell building timeline for a patient. The parameters involved are window length, grace period, minimum concurrent drugs and minimum spell length, as well as the list of ICD-10 ADEs and target NDC drugs.

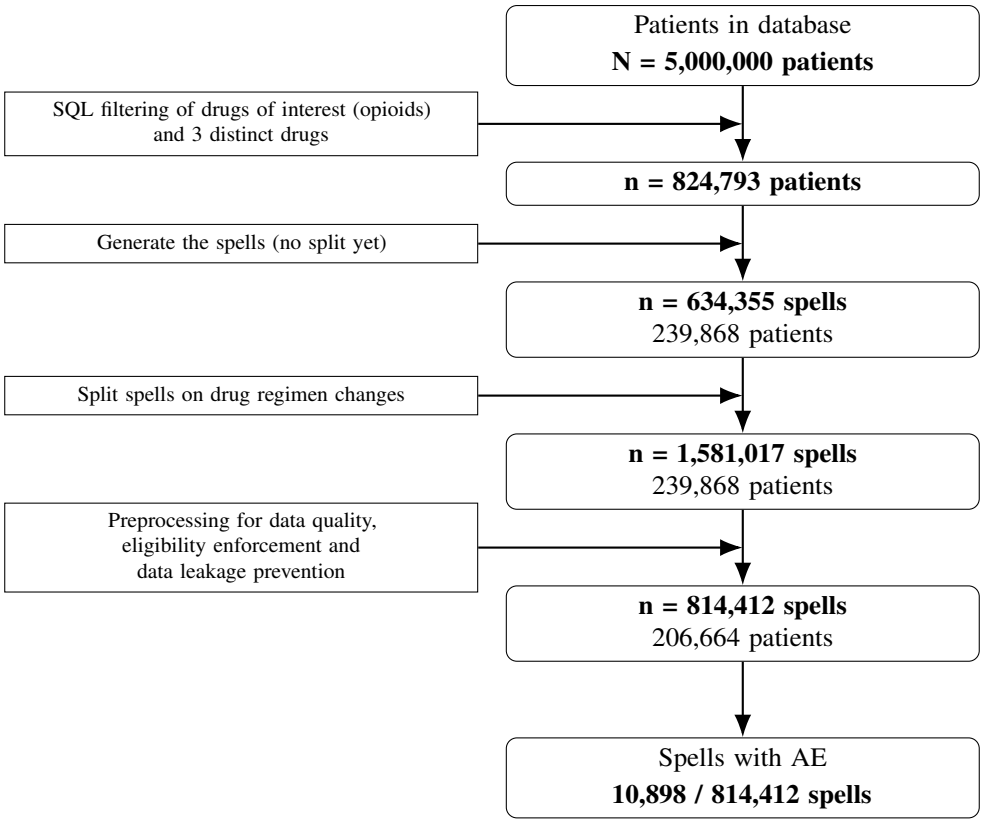
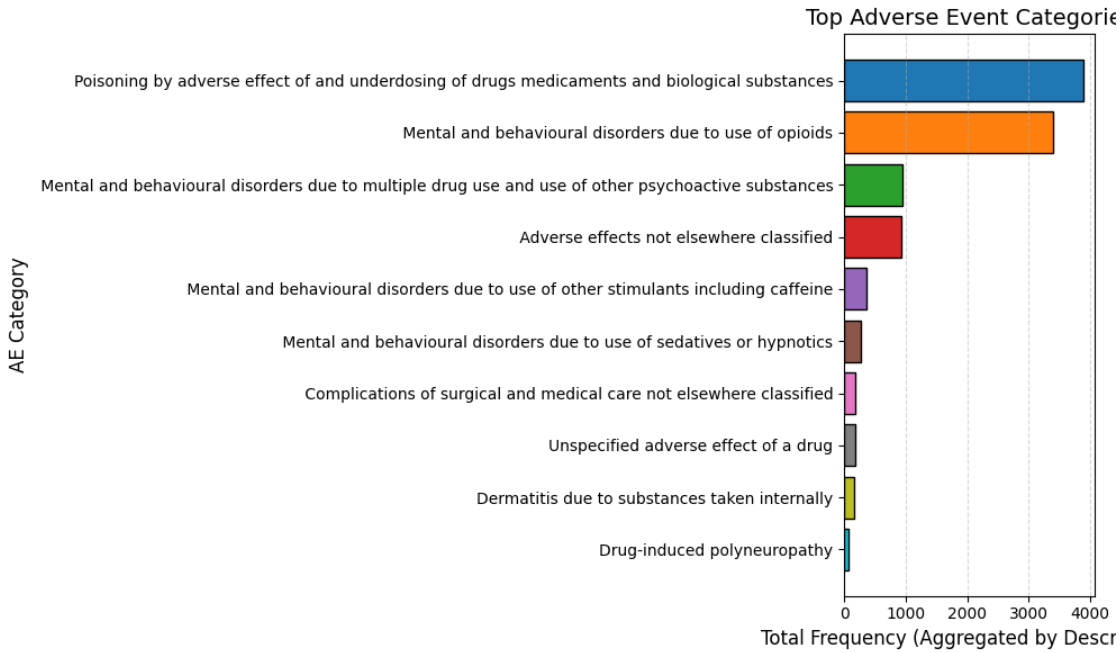


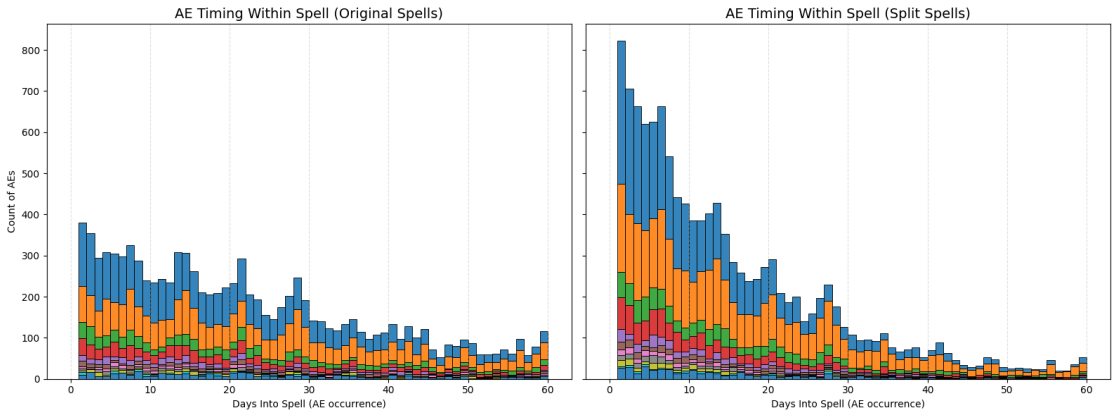
Figure 2: CONSORT-style diagram of the data extraction and processing steps, starting from the 5M member Inovalon sample.

Table 1: Baseline characteristics of opioid spells with and without an adverse event (ADE) within 30 days.

	AE within 30d	No AE within 30d
N spells	10,898	803,514
N patients	10,146	205,221
Age, median [IQR]	55.0 [44.0, 62.0]	56.0 [46.0, 63.0]
Female, %	62.8	63.9
Male, %	37.2	36.1
Other/Unknown gender, %	0.0	0.0
White, %	38.1	33.6
Black or African American, %	13.3	12.6
Hispanic or Latino, %	5.6	6.2
Asian or Pacific Islander, %	0.9	1.2
Some Other Race, %	4.9	4.9
Unknown, %	37.3	41.6
Distinct ICD10 codes per spell, median [IQR]	24 [13, 38]	11 [0, 23]



(a) Top adverse event (ADE) categories, aggregated by ICD-10 description. Bar colors match the timing distributions in panel (b).



(b) Timing of ADE occurrence within polypharmacy spells, comparing original and split spell definitions.

Figure 3: Distribution and timing of adverse events (ADEs) in the opioid-involved polypharmacy cohort.

Table 2: Highest ae prop Opioid drug pairs (baseline 1.70e-02). Pvalue from chi-sq test and adjusted with Benjamini Hochberg. Drugs in ATC3 class codes. Filtered to top 20 significant rows with n_total >200

drug1	drug2	n_total	n_ae	n_no_ae	ae_prop	p_value	p_adj
L01E	N02A	1418	83	1335	5.85e-02	6.59e-37	1.26e-34
L01X	N02A	1106	52	1054	4.70e-02	4.56e-16	2.01e-14
L03A	N02A	937	42	895	4.48e-02	4.52e-12	1.22e-10
L01A	N02A	205	9	196	4.39e-02	3.59e-03	1.53e-02
H01B	N02A	251	11	240	4.38e-02	1.08e-03	5.52e-03
B05B	N02A	462	18	444	3.90e-02	1.71e-04	1.12e-03
A12C	N02A	398	14	384	3.52e-02	4.27e-03	1.78e-02
A07A	N02A	3693	124	3569	3.36e-02	2.07e-17	1.04e-15
N02A	V03A	5188	173	5015	3.33e-02	2.25e-23	2.10e-21
B03X	N02A	401	13	388	3.24e-02	1.51e-02	5.21e-02
C01E	N02A	1701	50	1651	2.94e-02	1.52e-05	1.31e-04
A03F	N02A	5180	151	5029	2.92e-02	4.81e-14	1.58e-12
A05A	N02A	1034	30	1004	2.90e-02	1.26e-03	6.36e-03
N02A	N04A	1661	48	1613	2.89e-02	3.96e-05	3.08e-04
A09A	N02A	2192	63	2129	2.87e-02	2.71e-06	2.81e-05
A04A	N02A	35115	993	34122	2.83e-02	9.08e-79	1.21e-75
A11D	N02A	780	22	758	2.82e-02	9.75e-03	3.57e-02
N02A	N05A	35681	1006	34675	2.82e-02	5.33e-79	1.07e-75
J04B	N02A	328	9	319	2.74e-02	1.51e-01	3.28e-01
A02A	N02A	2818	74	2744	2.63e-02	1.78e-05	1.50e-04

Table 3: Highest ae prop Opioid drug trios (baseline 1.70e-02). Pvalue from chi-sq test and adjusted with Benjamini Hochberg. Drugs in ATC3 class codes. Filtered to top 20 significant rows with n_total >200

drug1	drug2	drug3	n_total	n_ae	n_no_ae	ae_prop	p_value	p_adj
A04A	L01E	N02A	251	21	230	8.37e-02	9.94e-17	2.44e-14
A04A	A09A	N02A	311	24	287	7.72e-02	5.05e-17	1.27e-14
A04A	N02A	N05A	2676	193	2483	7.21e-02	2.10e-118	4.43e-114
C08D	J01E	N02A	249	17	232	6.83e-02	2.41e-10	1.61e-08
A02B	L01X	N02A	221	15	206	6.79e-02	3.91e-09	1.98e-07
A07A	N01B	N02A	229	15	214	6.55e-02	1.08e-08	4.69e-07
A07A	H02A	N02A	551	36	515	6.53e-02	1.11e-19	3.90e-17
A04A	M04A	N02A	451	29	422	6.43e-02	1.22e-15	2.31e-13
A04A	N02A	V03A	437	27	410	6.18e-02	9.21e-14	1.24e-11
L01E	N02A	N06A	300	18	282	6.00e-02	4.78e-09	2.33e-07
A07A	B05X	N02A	308	18	290	5.84e-02	1.06e-08	4.62e-07
A02B	L01E	N02A	364	21	343	5.77e-02	8.12e-10	4.79e-08
L01E	N02A	N02B	279	16	263	5.73e-02	1.32e-07	4.23e-06
A05A	N02A	N02B	233	13	220	5.58e-02	4.46e-06	8.74e-05
A10A	N02A	P01A	236	13	223	5.51e-02	5.77e-06	1.09e-04
B05X	N02A	P01A	200	11	189	5.50e-02	3.76e-05	5.41e-04
A03B	A04A	N02A	273	15	258	5.49e-02	9.74e-07	2.34e-05
A07A	A11C	N02A	201	11	190	5.47e-02	4.08e-05	5.81e-04
C01E	M03B	N02A	256	14	242	5.47e-02	2.69e-06	5.66e-05
A04A	A07D	N02A	605	33	572	5.45e-02	1.26e-13	1.63e-11

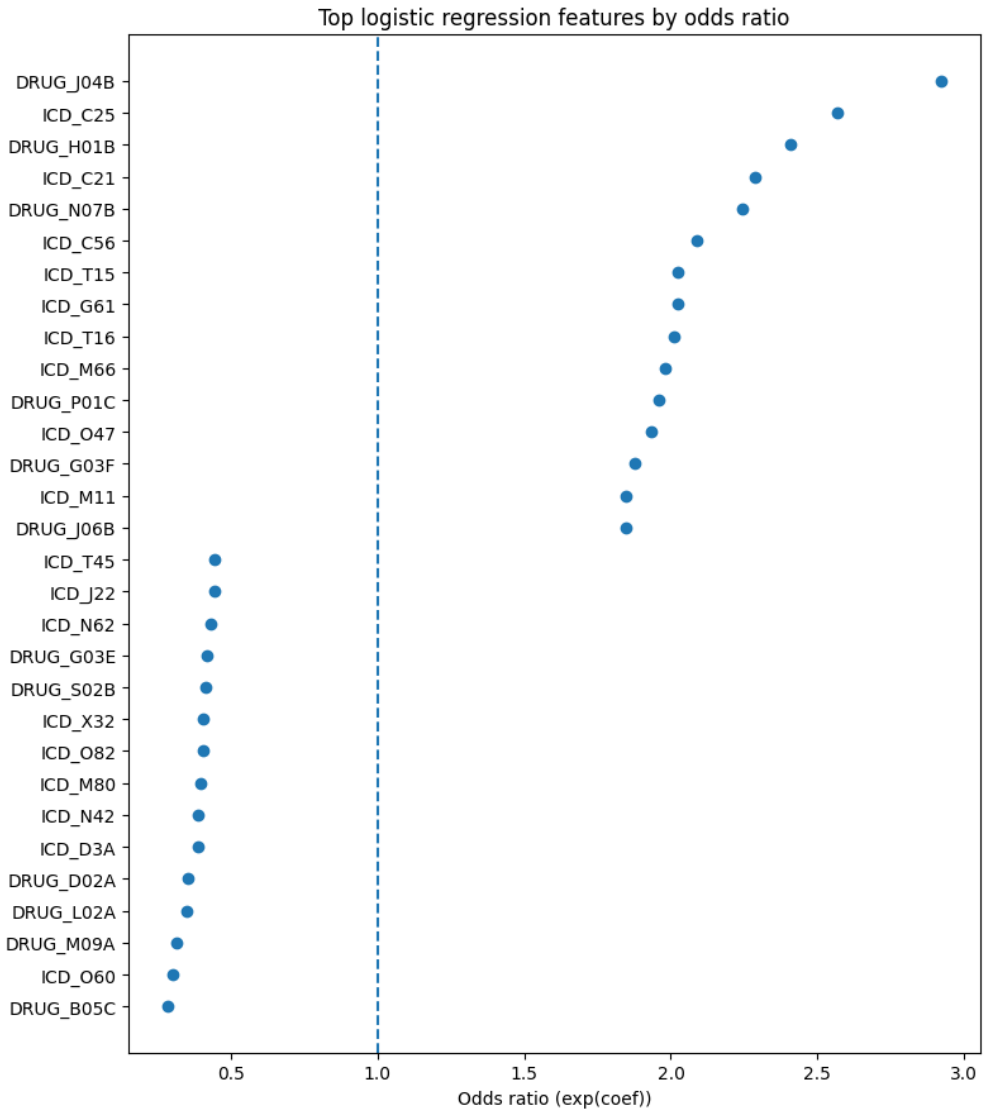
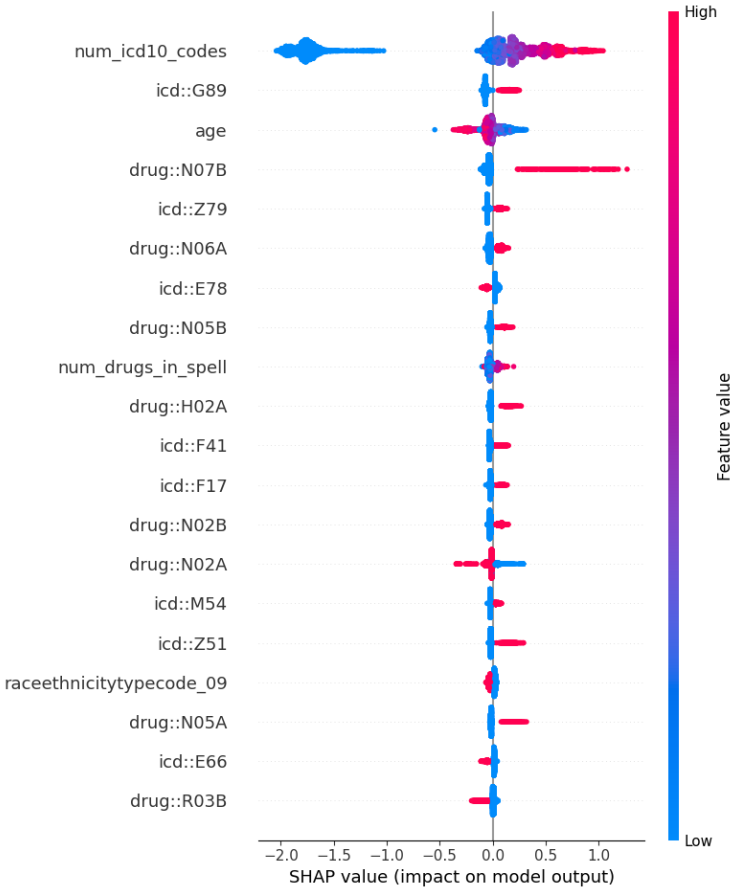
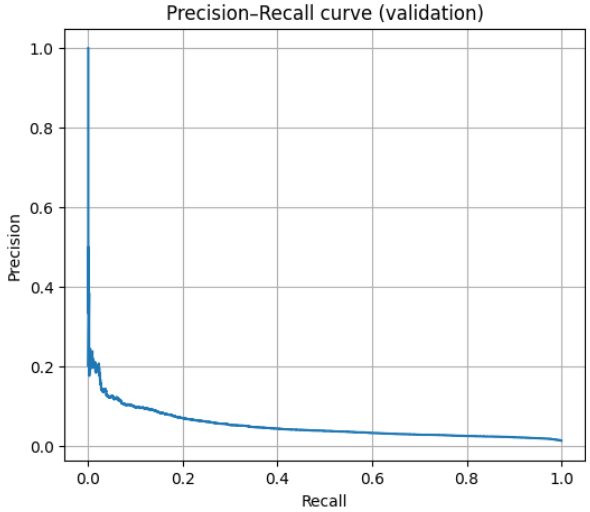


Figure 4: Odds ratios from the LogReg baseline analysis. Drugs are in ATC3 codes. The CIs could be built with bootstrapping.



(a) SHAP summary plot of feature contributions for the XGBoost model.



(b) Precision–recall curve of the best-performing XGBoost model.

Figure 5: Interpretability and performance plots for XGBoost model.

Methods	AUROC	AUPRC*	Precision @ top 0.1%	Precision @ top 5%
Logistic regression	0.7441	0.0441	16.0%	5.8%
XGBoost	**0.7672**	**0.0518**	20.3%	6.3%
MLP + MedTok	0.7186	0.0381	–	–

Table 4: Model performance on ADE prediction (ADE rate $\tilde{1}.3\%$).