Improving Drug Name Normalization in FAERS Using RxNorm and Approximate Matching

Caleb Stanberry, Dr. Qiang Guan, Zixu Wang

Choose Ohio First

KENT STATE UNIVERSITY

Introduction

- FAERS contains millions of drug names [1]
- Drug names are inconsistent (brand names, misspellings, formatting) [2]
- Standardization is required for pharmacovigilance [3]
- RxNorm provides a standardized drug vocabulary [4]

Example Drug Name Normalization

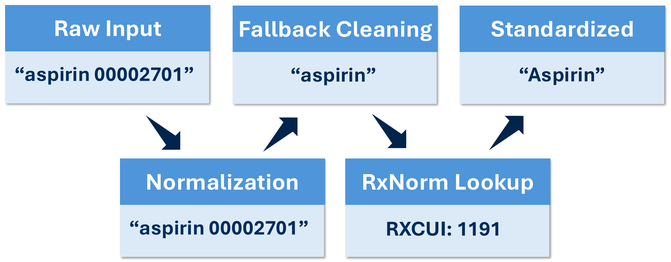

Figure 1. This example illustrates removal of numeric codes. Other common failure types included brand names, concatenated combination products, and formulation/device-specific suffixes.

Data

- Source: FDA Adverse Event Reporting System (FAERS), quarterly reports (2012–2023)
- Total drug records: 38.9 million
- Unique normalized drug names: 414,909

To improve efficiency and impact, analysis focused on the top 5,000 most frequent drug names:

- Covers **35.5 million records (91.3% of total data)**
- Captures majority of high-impact adverse event reports

Methods

- Deterministic preprocessing: Rule-based normalization (lowercasing, removal of punctuation and numeric artifacts, and whitespace normalization)
- RxNorm mapping (exact + approximateTerm): Staged lookup strategy using the RxNorm API, first attempting exact matches, followed by fallback cleaning and approximate matching (threshold ≥ 8.5)

Drug Name Normalization Pipeline

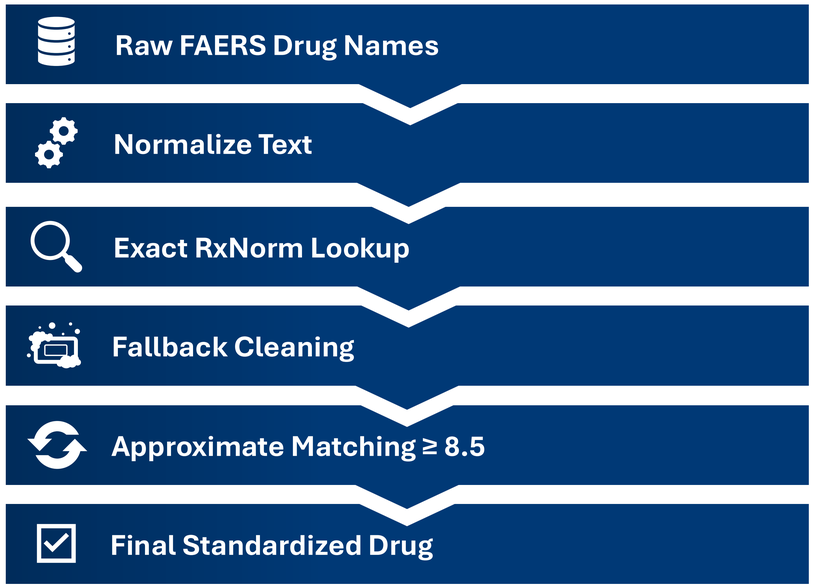

Figure 2. Deterministic pipeline for standardizing FAERS drug names using exact matching, rule-based cleaning, and approximate matching. This staged approach prioritizes precision while progressively recovering difficult-to-match drug names.

Results

Row Coverage by Method

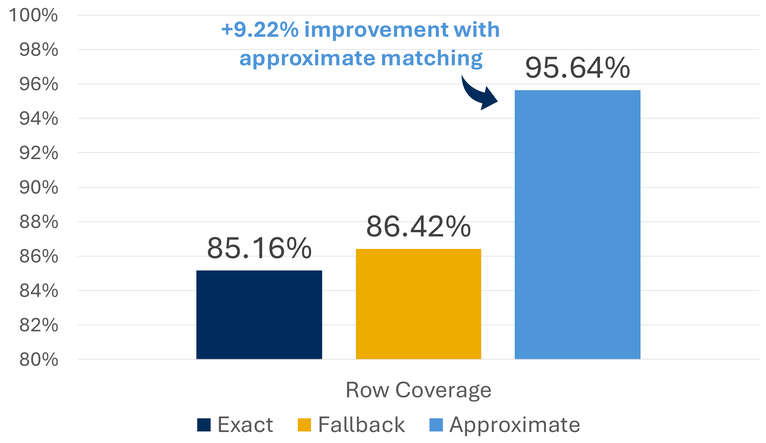

Figure 3. Comparison of row-level drug name coverage across exact, fallback, and approximate mapping methods. Approximate matching increases total coverage to 95.64%, demonstrating that many high-frequency drug names cannot be resolved through exact matching alone.

Method	% Names Mapped	Row Coverage
Exact RxNorm	56.98%	85.16%
+ Fallback	58.62%	86.42%
+ Approximate (≥ 8.5)	**84.20%**	**95.64%**

Table 1. Comparison of mapping performance across methods. While exact matching provides strong baseline coverage, fallback and approximate strategies are necessary to capture additional variation in real-world drug naming.

Conclusion

- Drug name normalization in FAERS requires more than exact matching due to widespread inconsistencies in real-world data.
- By combining deterministic preprocessing with approximate RxNorm matching, total row-level coverage increased from **85.16%** to **95.64%**.
- These results show that a staged, precision-first pipeline can effectively standardize large-scale pharmacovigilance data while preserving accuracy.
- Improved normalization enables more reliable downstream analysis of adverse drug events.

Future Work

Ingredient-level normalization
- Map drug products to active ingredients for more consistent aggregation of adverse events

Improved handling of complex drug names
- Address concatenated combination products and multi-ingredient formulations

Expansion beyond top 5,000 drugs
- Extend normalization to lower-frequency drug names

Integration into downstream analysis
- Apply standardized drug names to adverse event signal detection and modeling

References

1. U.S. Food and Drug Administration. (2023). FDA Adverse Event Reporting System (FAERS) Public Dashboard. https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html
2. Giunchi, V., Fusaroli, M., Hauben, M., Raschi, E., & Poluzzi, E. (2023). Challenges and Opportunities in Accessing and Analysing FAERS Data: A Call Towards a Collaborative Approach. *Drug safety*, *46*(10), 921–926. https://doi.org/10.1007/s40264-023-01345-w
3. Ogwo, M., Dhake, P., Schumaker, R., Dixit, R., & Veronin, M. (2020). A systematic approach to cleaning of drug name records data in the FAERS database: A case report. *International Journal of Big Data Management*, *1*(2), 105–118. https://doi.org/10.1504/IJBDM.2020.10034546
4. Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T., & Moore, R. (2011). Normalized names for clinical drugs: RxNorm at 6 years. *Journal of the American Medical Informatics Association : JAMIA*, *18*(4), 441–448. https://doi.org/10.1136/amiajnl-2011-000116