

Temporal Information Enhanced Association Rule Discovery and Evolution Pattern Identification in Complex Pharmaceutical Transaction Data

Haiyang Chen*

Ren'ai College of Tianjin University, Tianjin 301636, China

*Corresponding email: 15998867452@163.com

Abstract

Objective: To bring temporal information into association rule mining on complex pharmaceutical transaction data, so that co-prescribing rules and their evolution over time can both be identified. **Methods:** Transaction records are reorganized into time-stamped drug-set sequences at a selected Anatomical Therapeutic Chemical (ATC) abstraction level. Temporal association rules with interval constraints are extracted using a residual-leverage measure that shrinks toward the null hypothesis. Each rule is represented as a trajectory of support, confidence and leverage across sliding windows and clustered into five evolution modes: stable, emerging, decaying, periodic and abrupt. **Results:** On a provincial procurement and insurance settlement dataset, the method compresses the rule space while preserving coverage. It also exposes emerging polypharmacy combinations earlier than cross-sectional screening. **Conclusion:** It provides an interpretable instrument for rational drug use surveillance, chronic disease monitoring, drug interaction signal detection and insurance fund risk warning.

Keywords

Temporal association rules, Pharmaceutical transaction data, Evolution pattern identification, Trajectory clustering, Rational drug use

Introduction

Transaction records from centralized procurement platforms, hospital dispensing systems and insurance settlement systems form a densely time-stamped archive of population medication behaviour. This archive reveals not only which drugs are dispensed together. It also reveals the time point, the order and the interval of dispensing [1].

Association rule mining has been applied to medication data for two decades. Yet most implementations follow the static market-basket formulation in which a transaction is an unordered item set. This formulation cannot distinguish a concomitant prescription from a sequential one. It also cannot describe how rule strength changes over time. Preserving sequential structure, however, improves phenotype prediction by more than thirteen percent.

Three difficulties remain. The first is rule explosion, caused by the sparse and long-tailed distribution of

pharmaceutical products. The second is drift, because a cross-sectional snapshot cannot distinguish a stable structure from a transient one. The third is limited actionability, since a very large rule set can hardly be reviewed. This study contributes a temporal transaction model, an interval-constrained rule discovery procedure with a shrinkage-corrected residual-leverage measure, and a trajectory module assigning an evolution mode to each rule [2].

Clinical motivation reinforces the technical gap. Polypharmacy has become routine in managing hypertension, diabetes and other chronic conditions. Clinically relevant interaction risk rises steeply once several agents are dispensed together. Surveillance practice relies largely on prescribing review at a single encounter. It cannot observe whether two agents were supplied on the same day or weeks apart. It also cannot observe whether their co-occurrence is intensifying after

a procurement adjustment. Temporal information therefore carries direct operational meaning rather than statistical refinement.

The objective follows from this gap. Settlement records are reorganised into time-stamped drug-set sequences that retain the order and interval of dispensing. Co-prescribing rules are extracted whose validity is limited to an admissible interval. The result is compressed to a set a review panel can inspect within one working session. Each surviving rule receives an explicit account of how its strength changed across the observation period. In place of a single ranking, the output therefore describes a set of structures that emerge, persist, decay or shift abruptly.

The method therefore treats time as a first-class dimension of association analysis rather than a post-hoc annotation. Rules are discovered within admissible intervals, filtered by a shrinkage-corrected measure that suppresses spurious low-frequency patterns, and then tracked across sliding windows so that their evolution becomes part of the output itself. This design allows surveillance systems to distinguish durable co-prescribing structures from transient ones, and to flag combinations whose strength is rising before they become clinically visible. The resulting rule set is compact enough for expert review, yet rich enough to support early warning, insurance risk assessment and rational drug use monitoring.

Related work

Temporal association rule mining was established by research on unexpected temporal associations. That work proposed the residual-leverage measure and the Mining Unexpected Temporal Association Rules given the Consequent (MUTARC) algorithm for patterns of low frequency yet clinical surprise. Another line characterizes temporal association strength within a case as a continuous function. It compresses several hundred candidate rules to a reviewable scale. A configurable framework supports multiple time steps, explicit intervals and concept abstraction. It also couples rules with patient evolution patterns [3].

On medication data, sequential pattern mining predicts the next prescription at the drug class level from insurance claims. Combining it with care trajectory

clustering recovers homogeneous treatment sequences. For longitudinal records, temporal pattern discovery based on the observed-versus-expected contrast with shrinkage identified persistent, transient and periodic risk shapes. Association rule mining outperformed additive and multiplicative interaction models in sensitivity, at above ninety-nine percent. Two gaps remain: Few studies use transaction-type data, and rule discovery and evolution analysis are separate stages [4]. Despite these advances, three properties of transaction-type medication data remain insufficiently addressed. First, most temporal formulations presume a curated longitudinal cohort, whereas settlement records arrive as institutional transactions with uneven visit frequencies and no explicit episode boundary. Second, evaluation usually stops at predictive accuracy, leaving compression and reviewability unmeasured although they determine whether an output can be acted upon at all. Third, evolution analysis is commonly performed after rule discovery, on rules already fixed by a global threshold. Thus, a pattern whose average strength is modest yet whose trajectory rises monotonically never enters the candidate pool [5]. This study treats the three jointly: Windowed re-estimation feeds directly into clustering, so that trajectory shape, not average magnitude alone, decides which rules survive.

Methodology

Temporal transaction modelling

A transaction record is a tuple of institution, patient pseudonym, drug code, quantity and settlement date. Records are mapped onto the ATC classification and aggregated at a selected abstraction level to balance sparsity against actionability. For each patient pseudonym, records are ordered by settlement date and partitioned into consecutive windows of fixed length, so that drugs in one window are concomitant, drugs in different windows are sequential, and the distance between windows defines the interval [6].

This representation preserves both within-window co-prescribing and between-window temporal ordering. It also allows the same patient to contribute multiple windows without requiring a curated longitudinal cohort. By varying window length and abstraction level, the analyst can trade resolution against statistical support.

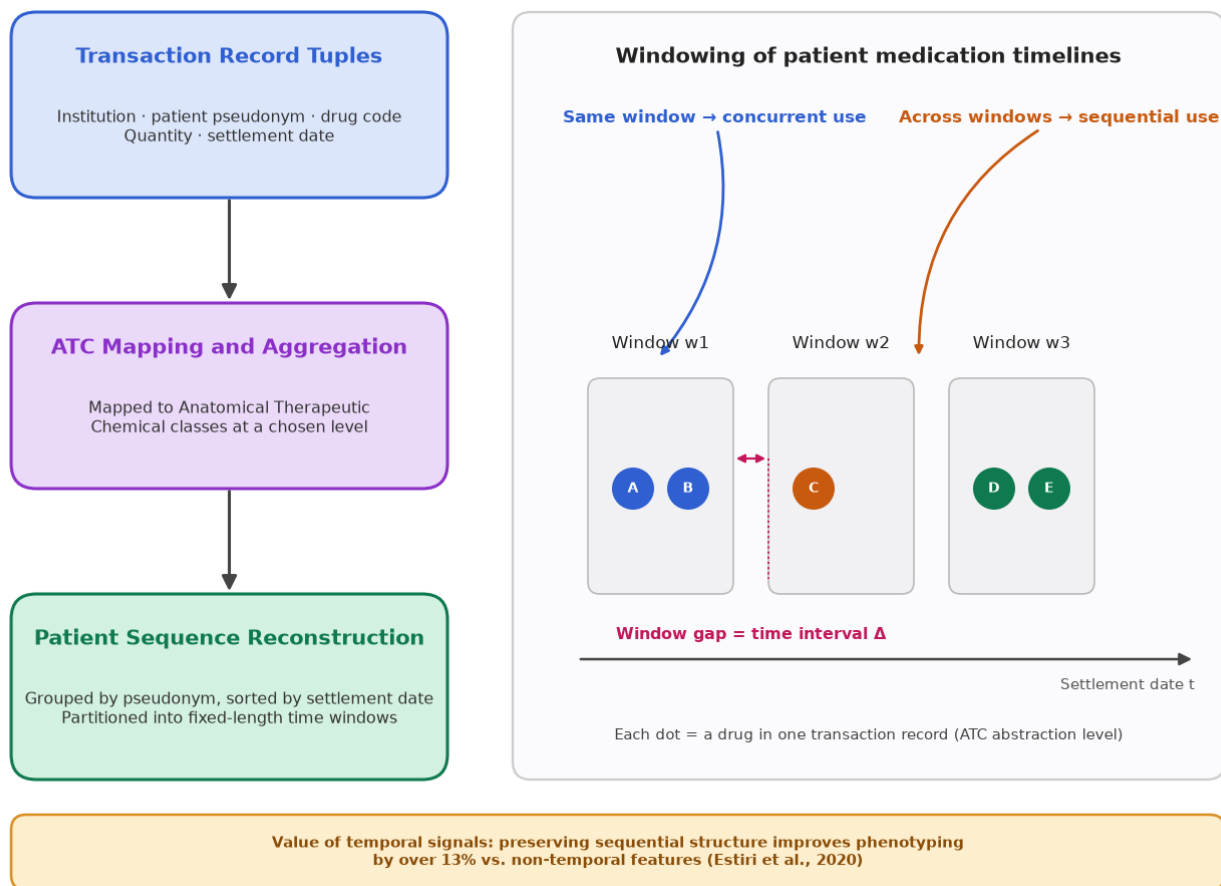


Figure 1. Temporal transaction modelling: windowed reconstruction from transaction record tuples to patient medication sequences.

Temporal association rule discovery

A temporal association rule is defined over two drug sets together with an admissible interval. It requires the antecedent to occur before the consequent within it. Support is counted only on windows satisfying the constraint, so concomitant prescriptions are not counted into sequential rules. Interest is measured by residual leverage, the difference between observed support and support expected under independence. Shrinkage toward the null is applied to suppress low-frequency spurious associations. Candidate rules are constrained by minimum support, maximum interval and abstraction level. Subsumed rules of comparable leverage are removed [7].

Procedural summary

The procedure is summarised here in operational terms. Records are mapped onto the ATC hierarchy. An abstraction level is selected. Level one yields broad classes with high support but limited clinical specificity. Level five is specific yet too sparse for reliable support estimation. The choice is made against an explicit

coverage target and reported alongside the results.

Windows then receive a length taken from the distribution of prescribing episodes rather than fixed arbitrarily. A minimum number of active windows per patient is required, so that no trajectory rests on isolated events. Candidate rules are enumerated under support and interval constraints. They are scored by shrinkage-corrected residual leverage. They are filtered for subsumption before surviving rules are re-estimated across sliding windows.

Two practical issues deserve emphasis. Sparse co-occurrence makes raw leverage unstable for rare combinations. Shrinkage addresses this by pulling estimates toward the null hypothesis in proportion to their low frequency. This is why the measure is preferred to raw confidence ranking. Overlapping windows induce serial dependence among successive estimates. Therefore, significance is stated at the cluster level rather than on individual windows. Computation is dominated by candidate enumeration and grows with abstraction granularity. Re-estimation over windows is

linear in the number of surviving rules and windows [8]. The rule set obtained over the full study period is re-estimated inside overlapping sliding windows, so each rule becomes a trajectory of support, confidence and leverage values. Trajectories are grouped by agglomerative hierarchical clustering with a correlation-

based distance, partitioning rules by how their strength changes rather than by how strong they are. Clusters are labelled stable, emerging, decaying, periodic and abrupt. Abrupt clusters are compared against an intervention indicator aligned to dated policy or supply events. This separates policy effects from normal variation.

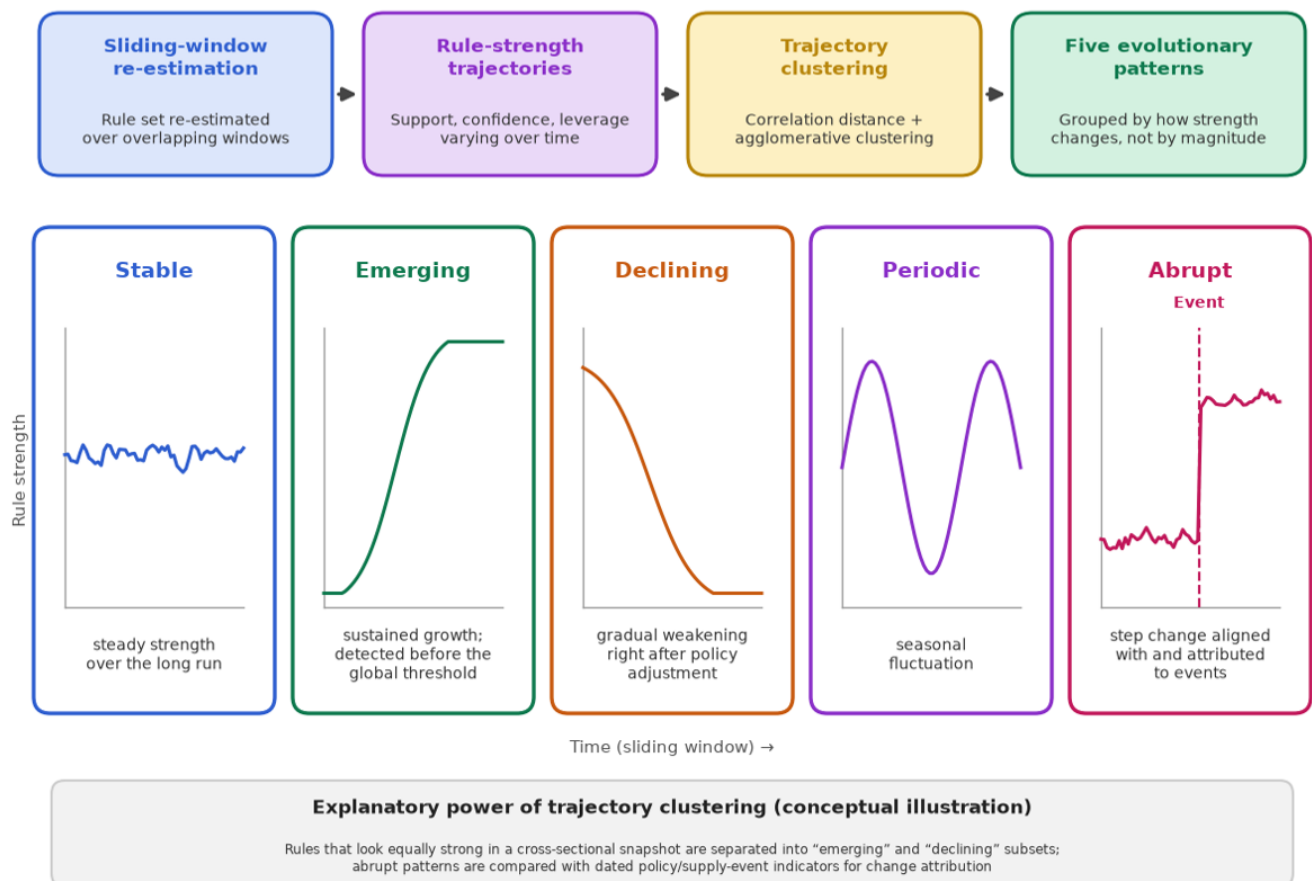


Figure 2. Evolution pattern identification procedure and five trajectory modes of rule strength (conceptual illustration).

Experimental design

The empirical analysis draws on a provincial procurement and insurance settlement dataset covering [start year] to [end year], with [N] transaction records, [M] drug products, and [K] institutions.

Records enter the analysis only where the drug code can be mapped to an ATC class. The settlement date must fall inside the observation period. The patient pseudonym must be stable across encounters. Extreme quantity values are winsorised at the ninety-ninth percentile. Bridging therapy after discharge, seasonal campaigns and short-term supply interruptions are deliberately retained. They are precisely the sources of periodicity and abrupt change that the evolution module is intended to detect. The effect is instead quantified by repeating the pipeline with those periods excluded. The

extract is de-identified at source and aggregated before analysis. No individual-level record leaves the secure environment [9].

Baselines are static Apriori at matched support thresholds, the Candidate Sequential Pattern Discovery Using Equivalence Classes (CSPADE) sequential pattern miner, the unexpected temporal association rule method and the functional temporal association method. Evaluation covers compression, measured by rules retained at comparable coverage. It also covers discriminative quality, measured by residual leverage and next-drug-class prediction accuracy. Usability is measured by expert ratings with Fleiss' kappa as the agreement statistic.

Parameters are set on a development period held out from [start year] and applied unchanged afterwards, so evaluation never tunes against the test period. Window

length is chosen from candidate values by the stability of cluster assignment under bootstrap resampling of patients. Support and leverage thresholds are chosen to meet a fixed rule budget agreed with the review panel. The maximum admissible interval is set at a high percentile of observed successive-dispensing gaps. All baselines receive an equal post-processing budget. Compression is compared only at matched coverage of dispensed volume, so that no method is rewarded merely for discarding rare but real combinations.

Results and analysis

Three behaviours are expected and testable. Interval constraints with abstraction-level aggregation should substantially reduce retained rules relative to static Apriori at equal coverage. Shrinkage-corrected residual leverage should rank clinically plausible combinations above rules driven by seasonal prescribing, and trajectory clustering should separate cross-sectionally strong rules into decaying and emerging subsets, the latter detectable several windows earlier [10].

Three robustness checks accompany the expected behaviours. Cluster assignment is resampled across patients to confirm that labels are not produced by a handful of large institutions. The interval constraint is varied to establish that rule compression does not simply follow from a wider window. Rules retained at coverage targets spanning the reported range are compared for overlap. This ensures that conclusions are not an artefact of one threshold. Expected failure modes are reported rather than concealed. Combinations dispensed almost exclusively inside one institution are flagged as local practice. Periodic patterns driven by seasonal campaigns retain their label while being excluded from the interaction-signal list. Abrupt shifts matching no dated event remain unattributed rather than being force-fitted to a policy indicator [11].

From a public health perspective the interpretable evolution modes are the main deliverable. An emerging cluster involving drugs with known interaction potential can be submitted to pharmacovigilance review. A decaying cluster appearing after a procurement adjustment serves as process evidence in policy evaluation. Clusters whose expenditure weight grows faster than their volume indicate fund risk. One limitation: Settlement records describe drugs prescribed

and reimbursed rather than actually taken, so case mix cannot be adjusted.

Operational deployment requires more than a ranked list. A reporting cycle that re-estimates trajectories quarterly produces a shortlist defined by change rather than by size. Review effort therefore concentrates on rules whose behaviour departed from the preceding period. Rules labelled stable are re-examined only when their window is refreshed. Interaction signals, procurement effects and expenditure risks are routed to different reviewers. The difference between successive cycles is archived, so a later disagreement can be traced to the window in which the change first appeared. Because every label derives from the same validation run, the pipeline can be re-executed retrospectively when coding practice or platform rules change. This keeps earlier conclusions comparable with current ones. Reviewers receive each label together with its trajectory plot, which keeps the judgement auditable.

Conclusion

This study proposes a temporal-information-enhanced method for association rule discovery and evolution pattern identification in complex pharmaceutical transaction data. Settlement records are reorganized into time-stamped drug-set sequences. Interval-constrained rules are extracted with a shrinkage-corrected residual-leverage measure. Each rule's strength trajectory is clustered into evolution modes. The method preserves the temporal signal that static market-basket formulations discard. The output is a compressed, clinically reviewable rule set. It includes an explicit account of how co-prescribing structures emerge, persist, decay or shift abruptly. Future work will extend the procedure to incremental updating.

Several limitations significantly bound the interpretation. Settlement data record what was prescribed and reimbursed, not what was taken, so adherence and case mix remain largely unobserved. Institutional coding practices vary considerably and unmapped products are discarded. The evolution labels describe association strength and carry no causal claim about the interventions they are aligned with. Three important directions follow: validating the emerging clusters against prospective pharmacovigilance records, replacing fixed windows with episode-aware

segmentation derived from the data, and moving re-estimation to an incremental form. In this form, a new reporting period updates existing trajectories instead of recomputing them from scratch.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

- [1] Ageno, A., Català, N., Pons, M. (2023) Acquisition of temporal patterns from electronic health records: an application to multimorbid patients. *BMC Medical Informatics and Decision Making*, 23(1), 189.
- [2] Feng, Y., Wang, K., Fan, J., Wu, X., Li, T., Yang, Z. (2024) Mindfulness intervention, homogeneous medical concept, and concentrated solution nursing for colorectal cancer patients: a retrospective study. *BMC Cancer*, 24(1), 1055.
- [3] Estri, H., Strasser, Z. H., Klann, J. G., McCoy, T. H., Jr., Waghlikar, K. B., Vasey, S., Castro, V. M., Murphy, M. K. E., Murphy, S. N. (2020) Transitive sequencing medical records for mining predictive and interpretable temporal representations. *Patterns*, 1(4), 1-11.
- [4] Assadi Shalmani, M., Khani, M., Taleban, A., Yi, Z., Fink, J. T., Weber, C. E., Luo, J. (2026) Sequential pattern transformer (SPT): a generative and interpretable framework for predicting disease trajectories. *Neural Computing and Applications*, 38(3), 28.
- [5] Wu, P., Nelson, S. D., Zhao, J., Stone Jr, C. A., Feng, Q., Chen, Q., Wei, W. Q. (2021) DDIWAS: High-throughput electronic health record-based screening of drug-drug interactions. *Journal of the American Medical Informatics Association*, 28(7), 1421-1430.
- [6] Mansour, A. M., Khazalah, F., Obeidat, M. A. (2022) Suspected adverse drug reaction detection using association rules mining and fuzzy sets. *Przegląd Elektrotechniczny*, 98.
- [7] Noguchi, Y., Ueno, A., Otsubo, M., Katsuno, H., Sugita, I., Kanematsu, Y., Yoshida, A., Esaki, H., Tachi, T., Teramachi, H. (2018) A new search method using association rule mining for drug-drug interaction based on spontaneous report system. *Frontiers in Pharmacology*, 9, 197.
- [8] Cai, R., Liu, M., Hu, Y., Melton, B. L., Matheny, M. E., Xu, H., Waitman, L. R. (2017) Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial Intelligence in Medicine*, 76, 7-15.
- [9] He, Y., Sun, J., Tan, X. (2025) Performance of apriori algorithm for detecting drug-drug interactions from spontaneous reporting systems. *Mathematics*, 13(11), 1710.
- [10] Hussein, A. H., Tarek, L., Maged, S., Safy, M. (2026) A comparative study of association rule mining algorithms for enhanced drug-drug interaction and adverse drug re-action discovery. *Journal of Al-Azhar University Engineering Sector*, 21(79), 1179-1189.
- [11] Flothow, A., Novelli, A., Sundmacher, L. (2023) Analytical methods for identifying sequences of utilization in health data: a scoping review. *BMC Medical Research Methodology*, 23(1), 212.