

Extracting and Visualizing Frequent Medical Instruction Patterns with Statistical Insights from Multi-Institutional Electronic Medical Record Data

Miwa Sugitani¹⁾, Ryosuke Matsuo¹⁾, Tomoyoshi Yamazaki¹⁾, Kenji Araki¹⁾,
Masato Oguchi¹⁾, Haruo Yokota²⁾, Hieu Hanh Le¹⁾

1) Ochanomizu University, Tokyo, Japan 2) Josai University, Tokyo, Japan

1. Research Background

- Widespread Electronic Medical Record (EMR) adoption enables large-scale medical data consolidation and secondary use.
- Analyzing medical data and clinical pathways improves treatment standardization and accuracy.
- Medical instructions vary by patient age and health condition.

2. Related Research

- Identified common closed patterns in single-institution EMRs, based on relative treatment dates and item counts [Y. Honda et al., DMAH, 2017].
- Compared two Sequence Variants (SVs) across multiple institutions, defining common, longest common, and merged SVs, and their calculation algorithm [Y. Li et al., DEXA, 2022].
- Analyzed temporal medical instruction patterns, revealing variable age, hospitalization, and test result influences in non-Hodgkin's lymphoma pathways [H.H Le et al. ACM Healthcare, 2023].

3. Problem

- Limited cross-institution/disease analysis.
- Insufficient statistical trend analysis.
- Basic patterns visualization only.

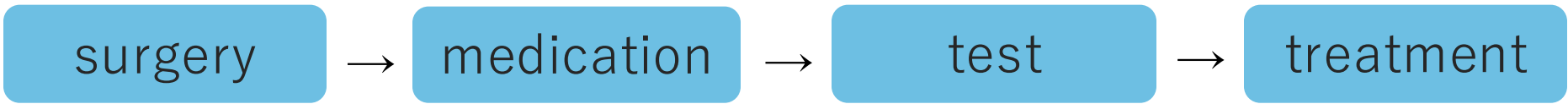
4. Research Purpose

- Extract disease-specific frequent medical patterns from integrated multi-institution EMRs.
- Derive statistical insights to compare frequent medical patterns.
- Develop a visualization tool displaying frequent pattern branching with statistical information.

5. Proposed Method

5.1. Create Medical Order Sequences

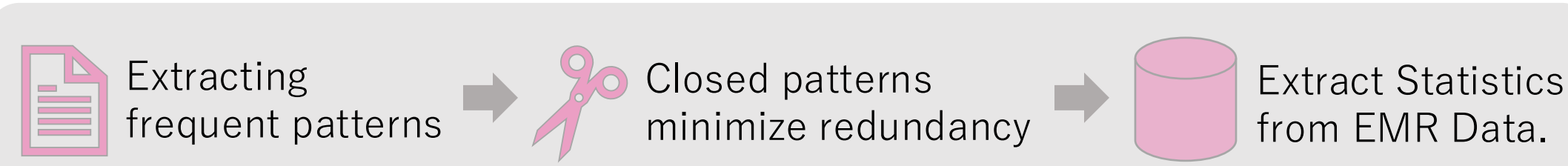
- Extract patient medical instructions, including surgery, medication, and examinations.
- Anchor instructions to Day 0 (surgery date), assigning and classifying them by days elapsed.
- Sort medical procedures performed on the same day in a defined order.



- Assign a sequence ID and store each sequence in a new data table.

5.2. Extract Frequent Medical Instruction Patterns with Statistical Information

- Frequent Pattern Extraction Using the T-PrefixSpan Algorithm [Uragaki et al., IEEE ICC, 2016]
 - Extracts ordered, temporal patterns, crucial for medical data.
- Reducing Redundant Data with Closed Patterns
 - Extract closed patterns to efficiently compact analysis results by eliminating redundancy.
- Extracting Statistical Information of Frequent Medical Instruction Patterns
 - Calculate key statistical information (mean, standard deviation, median, min/max) for frequent test patterns using patient data.

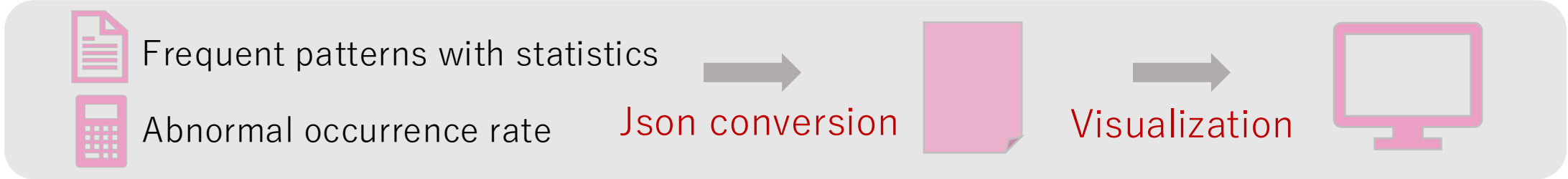


5.3. Calculate the Occurrence of Abnormal Test Results

- Extract test results by Sequence ID (SID), noting inclusion/exclusion in frequent medical order patterns.
- Count high and low abnormal values by day elapsed.
- Calculate for each test result: mean, standard deviation, 95% CI, low/high abnormal ratios, and N count.

5.4. Visualize Patterns and Results

- Convert the extracted data to json format
- Load the data into a visualization tool to visualize patterns of medical instructions for each disease and distribution of abnormal



6. Evaluation Experiment

Purpose

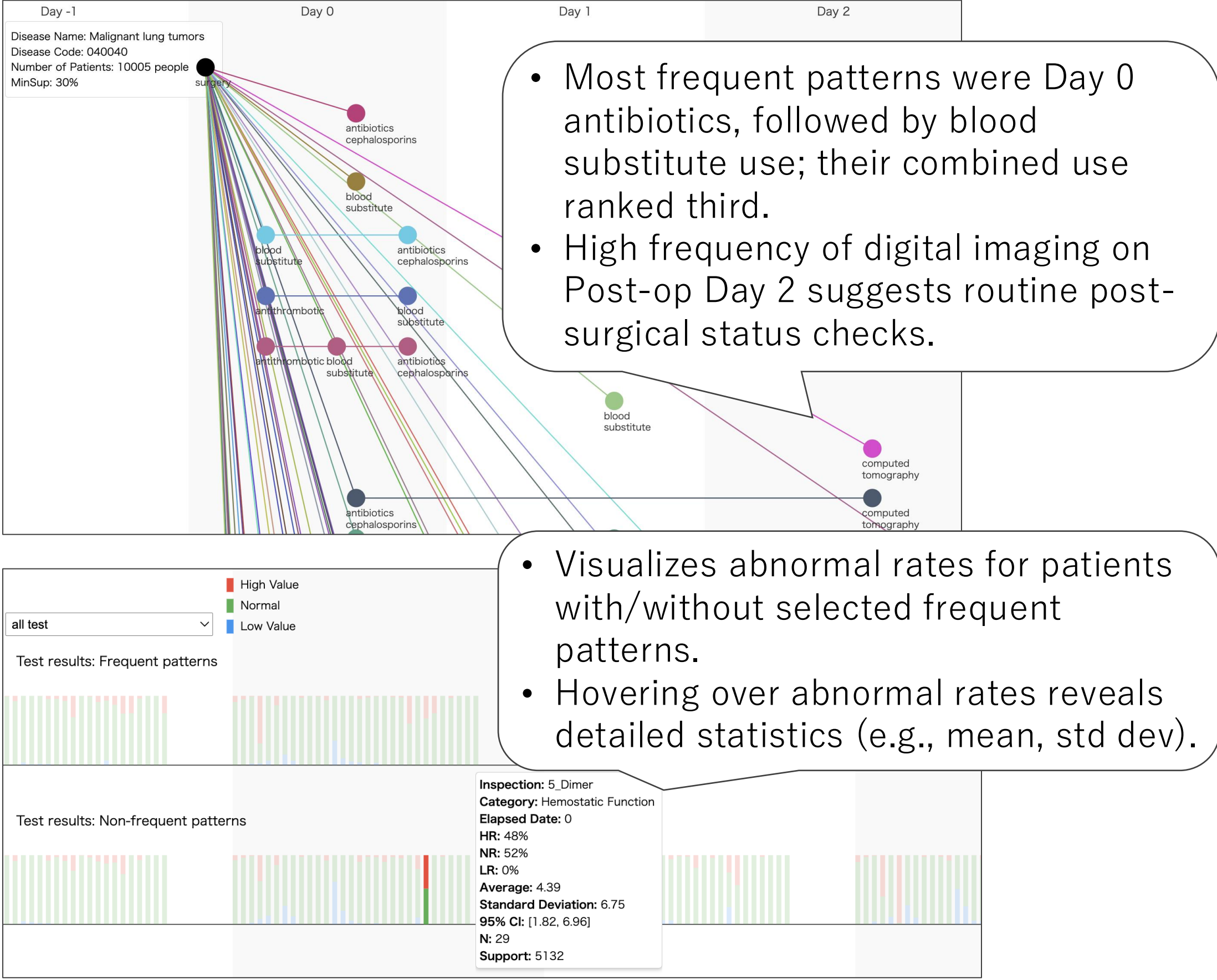
- Verify the extraction of useful patterns.

Experimental data

- Anonymized and standardized multi-institution EMR data, approved by the Ethics Committee of the Life Data Initiative.

Disease name	Patients
Malignant breast tumors	12,841
Angina, chronic ischemic heart disease	11,044
Malignant lung tumors	10,005
Bladder tumor	9,720
Malignant tumors of the stomach	5,961
Acute myocardial infarction	4,929
Malignant tumors of the liver and intrahepatic bile duct	2,016
Intervertebral disc degeneration, herniation	1,194
Malignant tumors of the cervix and uterine body	888
Total	58,598

Visualization results



7. Summary and Future Work

Summary

- Developed an application for visualizing frequent medical patterns, statistics, and abnormal test results' rates and statistics.
- Confirmed application effectiveness with real data.

Future Work

- Statistically analyze how patient characteristics (e.g., age, sex, comorbidities, lab results) influence frequent medical pattern branching.
- Compare patterns with clinical pathways, optimize extraction parameters for robust patterns, and verify medical validity.
- Develop a suggestion interface for healthcare professionals to reference relevant instruction patterns and tests.