Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230072

Data Sharing Platform for MIMIC-IV and MIMIC-ED Data Marts: Designing a Data Retrieving System Based on the Intra-Hospital Patient Transfer Pathway

Sharareh ROSTAM NIAKAN KALHORI[a,b 1], Thomas M. DESERNO[a], Jamal SOLEIMAN[a] and Shayan KASIRI HABIBABADI[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany*
[b] *Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran*
ORCiD ID: Sharareh Rostam Niakan Kalhori https://orcid.org/0000-0002-7577-1200
ORCiD ID: Thomas Deserno https://orcid.org/0000-0003-3492-4407
ORCiD ID: Jamal Soleiman https://orcid.org/0000-0001-7842-9366
ORCiD ID: Shayan Kasiri Habibabadi https://orcid.org/0000-0003-1871-790X

Abstract. Accessibility to high-quality historical data for patients in hospitals may facilitate related predictive model development and data analysis experiments. This study provides a design for a data-sharing platform based on all possible criteria for Medical Information Mart for Intensive Care (MIMIC) IV and Emergency MIMIC-ED. Tables containing columns of medical attributions and outcomes were studied by a team of 5 experts in Medical Informatics. They completely agreed about the columns connection using subject-id, HDM-id, and stay-id as foreign keys. The tables of two marts were considered in the intra-hospital patient transfer path with various outcomes. Using the constraints, queries were generated and applied to the backend of the platform. The suggested user interface was drawn to retrieve records based on various entry criteria and present the output in the frame of a dashboard or a graph. This design is a step toward platform development that is useful for studies aimed at patient trajectory analysis, medical outcome prediction, or studies that require heterogeneous data entries.

Keywords. Platform, MIMIC, design, retrieving, prediction model, SQL

1. Introduction

Retrospectively collected medical data provide the opportunity to improve patient care through algorithm development and knowledge discovery through modeling and outcome prediction. The model has higher quality in case of being developed with proper inputs in terms of enough quantity and dimensionality [1, 2]. Thus, data sharing for model development should provide the required records based on the defined problem [2, 3].

[1] Corresponding Author: Sharareh Rostam Niakan Kalhori, Muehlenpfordtstr. 23, 38106 Braunschweig, Germany. Contact details: +49 531 391-2125, Email: sharareh.niakankalhori@plri.de.

Despite the advances in patient data collection through electronic health records (EHR), registries, and self-care apps [2], data access remains a challenge, particularly concerning big data analysis. Sharing the whole big dataset when only a part of the data is required results in adverse consequences in terms of research ethical issues and waste of time for data understanding and preparation. The well-adjusted access to medical data based on the defined study questions may overcome multifaceted concerns. Platforms are a solution to narrow the data based on the set queries and criteria [2].

The MIMIC-IV and MIMIC-ED are currently shared via the Physionet website [4] in the frame of several separate CSV files [3, 5]. They contain the data of common records with specified ID from the emergency department (MIMIC-ED) to hospital wards and ICU with clinical details (MIMIC IV) providing the possibility of following each case's transfer across the hospital and the final outcome. Currently, data for cases at hospital departments and reports for radiology, laboratory, medication, clinical notes, vital signs in chart events, history of the disease, and demographic and clinical data are provided. The MIMI-IV is the medical information for over 40,000 patients admitted to intensive care units (ICU). The newer versions of the data have been even published with more features and volume of data[3].

Although the MIMIC-III database adopted a permissive access scheme that allowed for broad reuse of the data [5], there is no data-sharing platform to manage the records according to the queries. The mechanism of schema has already been used for the prediction of key patient outcomes such as mortality, clinical deterioration, and sepsis [6]; however, the data access for a record in two different marts with the same id, to follow the patients' pathway in hospital wards or discharge, remained a lack. Furthermore, learning the metadata of the MIMIC marts and manually extracting records from the mart prolonge the accessing process. Designing a platform is a step forward to having a tool for overcoming these limitations and retrieving the data for clinical and research purposes.

Due to the entity of available data in these marts with connected records of tables via subject-id starting from ED admission to ICU discharge, and to have a logical design for the sharing data platform, the patient intra-hospital patient transfer pathway is suggested. It starts from the emergency department (ED) where the patient refers or is transferred by ambulance for further care [1, 7]. Transfer from ED to inpatient wards is a common event, with over 12 million events annually [1]. Additionally, there are four million patients are admitted to ICU each year, either from hospital wards, or ED. Most of these patients transfer from ICU to a general ward (GW) [8]. In each point of care including ED, GW, or ICU, there are possible outcomes that three of them are covered in MIMIC-ED and MIMIC-IV including transfer or admission to the next station, discharge, and death. Hence, using the patient pathway structure may facilitate patient trajectory analysis and outcome prediction by retrieving the corresponding cases. This study aimed to design a system for a data-sharing platform for MIMIC-IV and MIMIC-ED based on an intra-hospital patient transfer pathway.

2. Material and Methods

In MIMIC-ED and MIMIC IV, tables are linked by identifiers which usually have the suffix 'ID'. For example, SUBJECT-ID refers to a unique patient, HADM-ID refers to a unique admission to the hospital, and ICU stay-ID refers to a unique admission to an

intensive care unit. They are unique across the patient transfer pathway and can be used to connect the columns of the marts' tables [1]. By joining Chartevent and given outcomes such as death via ID-items, it is possible to create the constraint and have a tailored new table. To fulfill the idea of designing a SQL-based platform, a technical expert team at PLRI of TU Braunschweig in Germany was created. After one year of working with these marts for experimental purposes, the technical team started designing a platform for easier and maximum usage of the available data as a preliminary step for system development. Seven focus group meetings, every 2 hours by 5 experts in the Medical Informatics and data engineering field were conducted. After identifying the IDs as primary and foreign keys of tables' columns, the queries based on SQL were studied [9]. That is, tables based on the mart structures are defined and connected. The experts agreed on the possible constraints for the backend of the platform; based on them, the features were considered for the front end of the platform.
With complete experts' concurrence regarding the front end of the platform's design, the following functions for the platform were considered:

- Storing the data in an SQL database as it is more suitable for our use case (dynamic search),
- Using the subject-id and HDM-id as foreign keys to relate an attribute such as the vital sign in chart time while staying in a specific department with a given outcome
- Using SQL query to get the relative information based on the search criteria
- Adding the option of exporting the results to a new CSV file
- The platform should contain a graph and dashboard section to visualize the data.

Experts completely agreed on these functions to be more customized according to the patient transfer pathway in hospitals in the developing step.

3. Results

The result of the first step of checking the common data elements in MIMIC-ED and MIMIC IV is presented in table 1. These columns of the tables in two marts could be connected via the subject-id, HDM-id, and ICU stay-id of the tables. To design the backend of the platform, the information schemas were depicted. Based on table 1 and the revealed platform functions, the frontend was designed, shown in figure 2.

Table 1. The available tables of marts containing columns of medical attributions and outcomes

	Tables	Example Columns	MIMIC-ED	MIMIC-IV
Tables of Attributes	Demographic	Age, sex, Education, race	✓	
	Administration	Billing, DRG, Services	✓	
	Clinical data	Diagnosis-ICD Procedures-ICD	✓	
	Measurements	Chart events, Procedure-events, Microbiology, Lab-events, Input-output event	✓	✓
	Medications	Prescriptions medication intake	✓	
	Clinical Notes	Radiology	✓	
Outcomes	ED Admission, ED transfer, ED discharge, ED death			✓
	Ward Amission, ward transfer, ward discharge, ward death			✓
	ICU admission, ICU stay, ICU death, ICU discharged		✓	✓

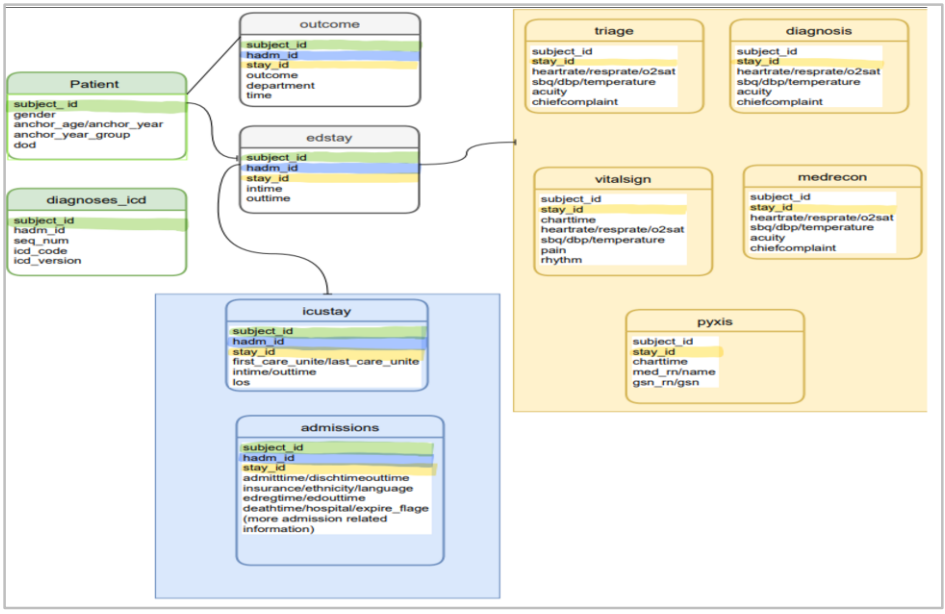

Figure 1. The example of an information schema of the backend of the platform.

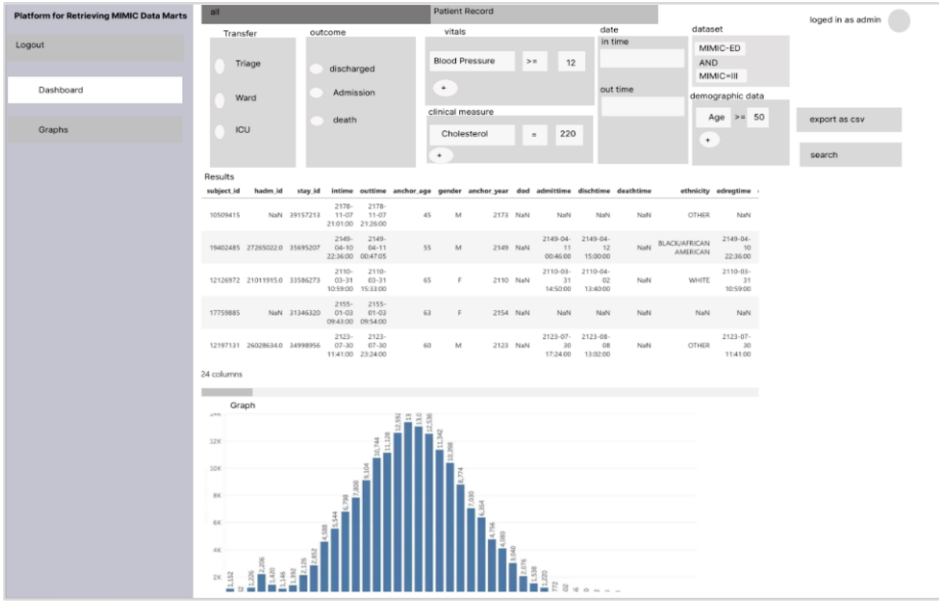

Figure 2. The user interface (UI) of the designed platform includes various data-extracting features including patient transfer, vital signs, selection of the Mart/s, date, demographic, and the given outcome. The extracted data will be provided in CSV and be shown in the frame of a graph in the dashboard section too.

4. Conclusion

According to the structure of MIMIC-IV and MIMIC-ED composed of the data of patients in the ED, triage, and admitted in the hospital (wards and ICU), designing a platform based on patient flow may be a solution for easy and quick data sharing. It is useful when the amount of data is continuously increasing. As Figure 1 shows there are several IDs presented in the same color that could connect data elements from different tables in the marts. These connections could be used for the criteria creation. As an example to get the records of a patient who died in the ICU with blood pressure greater than 140, the following steps should be done by the system:

- Getting the subject-id and stay-id of patient A to search the ICU table of MIMIC-IV
- Creating the constraints of death with BP>14 in ICU for the subject- and stay-id
- Searching the outcome column for the patients with the subject-id of patient A.
- Searching the death cases in the outcome table and picking the related record
- Using the constraints to bring up the required data for patient A in CSV.

This will be used to create all possible criteria to develop the data retrieving platform toward efficient data management and supporting researchers with heterogeneous data requirements. Examples of these studies could be titles such as prediction the ICU length of stay for patients transferred from ED with comorbidity and unstable vital signs, or emergency triage tool development to estimate the risk of ICU transfer for elderly patients affected with diabetes type II, or the trajectory prediction after hospital admission for pregnant women with unstable BP. Designing a platform as an electronic tool might be an essential need for easing the data analysis and knowledge discovery profession. However, it may face challenges regarding the limitation in data accuracy, not organized timing, and lab event data based on the unit. In the next steps, the research team has the plan for the system development and evaluation to facilitate the marts' usage.

References

[1] Xie F, Zhou J, Lee JW, Tan M, Li S, Rajnthern LS, et al. Benchmarking Predictive Risk Models for Emergency Departments with Large Public Electronic Health Records. arXiv preprint arXiv:211111017. 2021.

[2] Shortliffe EH, Chiang MF. Biomedical informatics: The science and the pragmatics. Biomedical informatics: Springer; 2021. p. 3-44.

[3] Gupta M, Gallamoza B, Cutrona N, Dhakal P, Poulain R, Beheshti R. An Extensive Data Processing Pipeline for MIMIC-IV. arXiv preprint arXiv:220413841. 2022.

[4] PhysioNet. Research Resource for Complex Physiologic Signals USA: National Institutes of Health (NIH); 1999 [cited 2023].

[5] Alistair L, Johnson E, Pollard T. Data descriptor: MIMIC-III a freely accessible critical care database. Thromb Haemost. 2016;76(2):258-62.

[6] Xia J, Pan S, Zhu M, Cai G, Yan M, Su Q, et al. A long short-term memory ensemble approach for improving the outcome prediction in intensive care unit. Computational and mathematical methods in medicine. 2019;2019.

[7] Leviner S. Patient flow within hospitals: A conceptual model. Nursing Science Quarterly. 2020;33(1):29-34.

[8] Al Owad A, Samaranayake P, Karim A, Ahsan KB. An integrated lean methodology for improving patient flow in an emergency department–case study of a Saudi Arabian hospital. Production Planning & Control. 2018;29(13):1058-81.

[9] Vathy-Fogarassy Á, Hugyák T. Uniform data access platform for SQL and NoSQL database systems. Information Systems. 2017;69:93-105.