*Challenges of Trustable AI and Added-Value on Health*
*B. Séroussi et al. (Eds.)*


*doi:10.3233/SHTI220522*

# MIMIC-IV as a Clinical Data Schema

Niklas GIESA[a1], Patrick HEEREN[a], Sophie KLOPFENSTEIN[a,b] Anne FLINT[a], Louis AGHA-MIR-SALIM[a], Akira PONCETTE[a], Felix BALZER[a], Sebastian BOIE[a]
[a] *Institute of Medical Informatics, Charité – Universitätsmedizin Berlin*
[b] *Berlin Institute of Health, Charité – Universitätsmedizin Berlin*

**Abstract.** Routinely collected electronic health records (EHR) in clinical information systems (CIS) are often heterogeneous, have inconsistent data formats and lack of documentation. We use the well-known open-source database schema of MIMIC-IV to address this issue aiming to support collaborative secondary analysis. Over 154 million data records from a German ICU have already been mapped and inserted into the schema successfully. However, discrepancies between the German and US health systems as well as specifics in our clinical source data hinder the direct translation to MIMIC. Evaluating and improving mapping completeness is part of the ongoing research.

**Keywords.** MIMIC, Electronic Health Records, Critical Care, Clinical Database

## 1. Introduction

The demand for exchanging standardized Electronic Health Records (EHR) increases with the maturity level of digital innovations in healthcare [1]. To develop and deploy clinical applications, EHR must be stored in a data model which mitigates issues like inconsistencies, format variety or incomplete documentation [1, 2]. The Medical Information Mart for Intensive Care (MIMIC) is a widely used clinical database providing an open-accessible well-documented data schema [3, 4]. Entities for patient characteristics, laboratory values, high-frequency measurements, and billing information are included [4]. Due to its lightweight design, we use MIMIC as a schema to store EHR from ICUs of a German university hospital enabling collaborative secondary analysis. Although previous work used MIMIC for data translation, no known study tried to map data directly from online clinical information systems (CIS) to the schema [5, 6].

## 2. Methods

The most recently published fourth version of MIMIC was chosen for mapping data fields from clinical MSSQL databases. MariaDB was selected to construct tables by executing the open accessible data definition language (DDL) scripts to create MIMIC-IV [4]. Extract, transform, load (ETL) processes run on Python 3.7.6 and execute SQL statements via the ODBC engine JayDeBe against the source and the target tables. Ethics approval was obtained by an independent ethics committee [7].

[1] Niklas Giesa, Institute of Medical Informatics, Charité – Universitätsmedizin Berlin; Charitéplatz 1 10117 Berlin, E-mail: niklas.giesa@charite.de

## 3. Results

Up to now, the MIMIC replica includes 37,460 patients with 54,949 ICU stays from March 2019 to October 2021. Eleven tables have been populated with data successfully incorporating more than 154 million EHR. 95% of data fields from the tables *admissions*, *patients* and *transfers* – storing information about the hospital stay – could be mapped to the schema. Mapping completeness ranged from 30 to 70% for *chartevents*, *inputevents*, *procedureevents* and *labevents* holding routinely collected high volatile data. 90% of billing information – stored in *diagnosis_icd* and *procedures_icd* – were transferred to the MIMIC schema; adaptions were necessary because of differences between German and US healthcare systems. 71% of data columns from *d_items* and *d_labitems* could be mapped. In total, 60% (78 out of 129) of data fields have been used for the study.

## 4. Discussion

Our results illuminate that the openly accessible MIMIC structure can be used to store data from German clinical source systems. However, the transferability of provided columns varies a lot; Paris et. al. state similar issues causing 16% to 80% loss of fields per table [5]. Furthermore, Maier et al. emphasize the lack of mapping concepts for German billing codes, hindering data translation in our study as well [6]. Besides, CIS – designed for daily clinical care – exhibit poor data quality and integrity [1, 2]. In order to provide a curated and interoperable database, further published approaches regarding clinical mapping, data validation and standardization must be evaluated [1, 2, 5, 8].

## 5. Conclusion

When keeping specifics of source data and health system discrepancies in mind, the schema of MIMIC can be used to efficiently store EHR extracted from German ICUs.

## References

[1] Lehne M, Sass J, Essenwanger A, Schepers J, et al. (2019). Why digital medicine depends on interoperability. NPJ digital medicine, 2(1), 1-5.
[2] Nissen F, Quint JK, Morales DR, et al. (2019). How to validate a diagnosis recorded in electronic health records. Breathe, 15(1), 64-68.
[3] Johnson AE, Pollard, TJ, Shen L, et al. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, 3(1), 1-9.
[4] https://mimic.mit.edu/docs/iv/, accessed 15.12.2021 01:09 Berlin, Germany
[5] Paris N, Parrot A (2020). MIMIC in the OMOP Common Data Model. medRxiv, 1-3
[6] Ulrich H, Behrend P, Wiedekopf J, et al. (2021). Hands on the Medical Informatics Initiative Core Data Set—Lessons Learned from Converting the MIMIC-IV. In German Medical Data Sciences 2021, (pp. 119-126). IOS Press. P. 1-5
[7] Ethics vote no. EA1/127/18, chairperson: Prof. R. Morgenstern, ethics committee: Campus Charité - Mitte
[8] Maier C, Lang L, Storf H, et al. (2018). Towards implementation of OMOP in a German university hospital consortium. Applied clinical informatics, 9(01), 054-061