

Database

Credentialed Access

# MIMIC-IV-ED

Alistair Johnson , Lucas Bulgarelli , Tom Pollard , Leo Anthony Celi , Roger Mark , Steven Horng 

Published: Jan. 5, 2023. Version: 2.2

**When using this resource, please cite:** [\(show more options\)](#)

Johnson, A., Bulgarelli, L., Pollard, T., Celi, L. A., Mark, R., & Horng, S. (2023). MIMIC-IV-ED (version 2.2). *PhysioNet*. <https://doi.org/10.13026/5ntk-km72>.

**Please include the standard citation for PhysioNet:** [\(show more options\)](#)

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.

## Abstract

MIMIC-IV-ED is a large, freely available database of emergency department (ED) admissions at the Beth Israel Deaconess Medical Center between 2011 and 2019. The database contains ~425,000 ED stays. Vital signs, triage information, medication reconciliation, medication administration, and discharge diagnoses are available. All data are deidentified to comply with the Health Information Portability and Accountability Act (HIPAA) Safe Harbor provision. MIMIC-IV-ED is intended to support a diverse range of education initiatives and research studies.

## Background






The emergency department (ED) is a high demand environment where patients are assessed and triaged for further care. ED patients compose a heterogenous cohort with severity ranging from mild abrasions to life-threatening cardiac complications. The ED is fundamentally a resource limited environment where the most important resource available, human attention, is rationed to maximize positive patient outcomes. Recent advances in algorithmic approaches present an exciting opportunity for improving the quality of care delivered in the ED. A prerequisite to data-driven analyses are sufficiently large datasets, and broad data accessibility enables reproducibility of research. MIMIC-IV-ED is intended to support data analysis in emergency care by providing a large database of admissions to an ED at a tertiary academic medical center in Boston, MA. It is a module of MIMIC-IV, meaning the information contained within MIMIC-IV-ED is linkable to information in MIMIC-IV [1].

## Methods

Data were extracted from the Beth Israel Deaconess Medical Center (BIDMC) ED in eXtended Markup Language (XML), and subsequently converted from XML into a denormalized relational database designed to simplify analysis. All data were deidentified to comply with the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provision [2]. Patient identifiers were replaced with randomized surrogates. Three deidentified patient identifiers are present in the dataset: `subject_id`, `hadm_id`, and `stay_id`. All three of these identifiers were generated in concordance with MIMIC-IV and MIMIC-CXR, allowing linkage of these datasets using one or more of the aforementioned identifiers [1,3-5]. Dates were shifted to a random time occurring between 2100 - 2200 on a patient-specific basis. Date shifts were consistently applied for a single `subject_id`, and all times associated with a single `subject_id` are temporally consistent and reflect the true order of events. Conversely, distinct `subject_id` which have data overlapping in time were not necessarily present

Contents

Share



Access

**Access Policy:**  
Only credentialed users who sign the DUA can access the files.

**License (for files):**  
[PhysioNet Credentialed Health Data License 1.5.0](#)

**Data Use Agreement:**  
[PhysioNet Credentialed Health Data Use Agreement 1.5.0](#)

**Required training:**  
[CITI Data or Specimens Only Research](#)

Discovery

**DOI (version 2.2):**  
<https://doi.org/10.13026/5ntk-km72>

**DOI (latest version):**  
<https://doi.org/10.13026/1cjn-2370>

**Topics:**

emergency department

ed

emergency

mimic-iv

electronic health record

mimic

**Project Website:**  
<https://mimic.mit.edu>

Corresponding Author

You must be logged in to view the contact information.

in the ED at the same time. Finally, free-text fields were processed using a hybrid deidentification algorithm, and PHI entities detected were replaced with three underscores ('\_\_\_') [6].

A schema comprising of six tables was created. The *edstays* table was created to track patient admission and discharge from the ED for a single patient stay as identified by *stay\_id*. Five data tables store information documented during the patient's stay: *diagnosis*, *medrecon*, *pyxis*, *triage*, and *vitalsign*. Tables are named to reflect the data within or its provenance. While a core aim of MIMIC-IV-ED is to provide real world clinical data for research purposes, and as such limit the amount of preprocessing performed prior to data release, a number of data cleaning steps were necessary during transformation. Observations were deduplicated upon insertion using the table specific primary key. The primary key was a combination of *stay\_id*, *charttime* if present, and additional attribute columns as appropriate (e.g. the *name* column in *pyxis*). For deidentification purposes, a regular expression was used to retain only numeric vital signs in the *triage* and *vitalsign* tables. Observations more than one year outside of the ED stay - usually occurring due to typographical errors in the charted time - were removed.

## Data Description

MIMIC-IV-ED is composed of a single patient tracking table, *edstays*, and five data tables: *diagnosis*, *medrecon*, *pyxis*, *triage*, and *vitalsign*.

### edstays

Patient stays are tracked in the *edstays* table. Each row of the *edstays* table has a unique *stay\_id*, which represents a unique patient stay in the ED. The *edstays* table contains the following columns: *subject\_id*, *hadm\_id*, *stay\_id*, *intime*, *outtime*, *gender*, *race*, *arrival\_transport*, and *disposition*. The *intime* indicates the time at which the patient was admitted to the ED, and the *outtime* indicates the time at which the patient was discharged from the ED. If the patient was admitted to the hospital following their ED stay, the *hadm\_id* column will be populated with an identifier representing their hospital stay. *hadm\_id* can be linked with the *hadm\_id* in MIMIC-IV to obtain further detail about the patient's hospital stay. Each individual is assigned a unique *subject\_id*, and patients with multiple ED stays will have the same *subject\_id* across stays in the *edstays* table. Patient demographics including race and gender are provided in the respective columns. The mechanism of patient admission is provided in *arrival\_transport*, and is coded into one five values: AMBULANCE, HELICOPTER, WALK IN, UNKNOWN, or OTHER. Patient discharge location is coded in *disposition*, and is one of eight values: ADMITTED, ELOPED, EXPIRED, HOME, LEFT AGAINST MEDICAL ADVICE, LEFT WITHOUT BEING SEEN, TRANSFER, and OTHER.

Note that *subject\_id* can be used to link MIMIC-IV-ED with MIMIC-IV to obtain additional information regarding individuals, e.g. age. *subject\_id* can also be linked with the PatientID DICOM attribute in MIMIC-CXR to obtain chest x-rays for patients if they are available [3].

### diagnosis

The *diagnosis* table provides coded diagnoses for the patient in the International Classification of Diseases (ICD) Ninth or Tenth revision (ICD-9 or ICD-10). These diagnoses are determined by trained coders after discharge from the emergency department and are used for billing purposes. There are six columns in the *diagnosis* table: *subject\_id*, *stay\_id*, *seq\_num*, *icd\_code*, *icd\_version*, and *icd\_title*. A maximum of 9 ICD codes are available for a single stay. The *seq\_num* column provides a pseudo-order for the ICD codes, with a value of 1 usually indicating highest relevance and a value of 9 indicating least relevance. The *icd\_code* provides the coded representation of the diagnosis using the ICD ontology, the *icd\_version* column is either 9 or 10 indicating whether the ontology used is ICD-9 or ICD-10, and the *icd\_title* column provides the textual description of the ICD code.

It is important to note that the billed diagnoses in the *diagnosis* table are exclusively related to the patient's emergency department stay. If the patient is subsequently admitted to the hospital, they will have a separate set of billed diagnoses for their

#### Versions

[1.0](#) - June 3, 2021

[2.0](#) - June 27, 2022

[2.2](#) - Jan. 5, 2023

hospital stay, which are not recorded in this table. See the usage notes for details regarding linking MIMIC-IV-ED to MIMIC-IV, which would facilitate comparison of the billed ED diagnoses with billed hospital diagnoses.

## medrecon

The *medrecon* table provides medicine reconciliation for each patient, that is a list of the medications which the patient was taking prior to their ED stay. The *medrecon* table has nine columns: `subject_id`, `stay_id`, `charttime`, `name`, `gsn`, `ndc`, `etc_rn`, `etccode`, and `etcdescription`. The `charttime` provides the date and time at which the medicine reconciliation was documented. The `name` column provides a text description of the medicine, the `gsn` column provides the Generic Sequence Number (GSN), and the `ndc` column provides the National Drug Code (NDC). Note a `gsn` or an `ndc` of 0 indicates that the value is missing. Columns prefixed with `etc` provide an ontology for grouping together drugs of a similar class. Note that as a medicine can be classified in multiple groups in the ontology, there may be more than one row for a single medication. For example, the medication Adderal is (1) a CNS stimulant, (2) an Attention Deficit-Hyperactivity Therapy, and (3) a narcolepsy therapy. As a result, patients taking adderal prior to their admission will have three rows in the *medrecon* table, delineated by the sequential monotonically increasing integer `etc_rn`. The `etccode` provides the coded form of the ontology group, and the `etcdescription` provides the textual description of the ontology group.

## pyxis

The *pyxis* table provides dispensation information for medications provided by the BD Pyxis MedStation, an automated medication dispensing system present in the ED [7]. The *pyxis* table has nine columns: `subject_id`, `stay_id`, `charttime`, `med_rn`, `name`, `gsn_rn`, and `gsn`. The `charttime` provides the time at which the medication was dispensed. If multiple medications were dispensed at the same time, the `med_rn` column delineates these medications. The `name` column provides a textual description of the medication dispensed, and may additionally contain auxiliary information such as the formulation. The `gsn` column provides the Generic Sequence Number (GSN) if available, and `gsn_rn` delineates multiple GSN values associated with the same medication. Note that a `gsn` of 0 indicates that the GSN is missing. Not all medications are dispensed by the Pyxis MedStation, and as a result not all medications are recorded in the *pyxis* table. For example, large fluid volumes (such as those used for resuscitation) are not present in this table.

## triage

The *triage* table provide information collected from the patient at the time of triage. All patients who present to the ED are immediately triaged, a process which involves assessing their health status and ascertaining the reason for their visit. The *triage* table has eleven columns: `subject_id`, `stay_id`, `temperature`, `heartrate`, `resprate`, `o2sat`, `sbp`, `dbp`, `pain`, `acuity`, and `chiefcomplaint`. Vital signs collected at triage include patient temperate (`temperature`), heart rate (`heartrate`), respiratory rate (`resprate`), oxygen saturation (`o2sat`), systolic blood pressure (`sbp`), and diastolic blood pressure (`dbp`). Although vital signs can be documented as free-text, the deidentification approach retained only numeric vital signs. A patient reported pain level is available in the `pain` column. The `chiefcomplaint` is a free-text field which contains the patient’s reported reason for presenting to the ED. The `chiefcomplaint` field is usually a comma separated list of entries. PHI present in the `chiefcomplaint` field has been replaced by three underscores ("\_\_\_"). Based upon the triage assessment, the care provider will assign an integer level of severity (`acuity`), where 1 indicates the highest severity and 5 indicates the lowest severity.

## vitalsign

The *vitalsign* table contains aperiodic vital signs documented for patients during their stay. The *vitalsign* table has eleven columns: `subject_id`, `stay_id`, `charttime`, `temperature`, `heartrate`, `resprate`, `o2sat`, `sbp`, `dbp`, `rhythm`, and `pain`. Vital signs in the *vitalsign* table are similar to those collected in the *triage* table. The `rhythm` column additionally provides the hearth rhythm for the patient. The `charttime` provides the time at which the vital signs were recorded.

---



# Usage Notes

## Organization

MIMIC-IV-ED is organized in a star schema, best understood visually, in which a single table is at the center of the star and all other tables link to this central table using the same identifier. The *edstays* table provides admission and discharge times for each stay in MIMIC-IV-ED, uniquely referred to by the identifier *stay\_id*. All other tables may be linked to *edstays* through *stay\_id*, and most tables have more than one row per *stay\_id*.

MIMIC-IV-ED may be analyzed using any number of software programs, including relational database management systems. Code for loading MIMIC-IV-ED into PostgreSQL is provided in an open source repository [8,9]. The repository also contains code for deriving concepts, tutorials, data analysis notebooks, and acts as a forum for community discussion [8,9]. We further provide MIMIC-IV-ED natively in cloud based database services including Google BigQuery, allowing immediate use of the dataset for credentialed investigators.

## Data Linkage

MIMIC-IV-ED is usable as a standalone research database, but may also be linked to MIMIC-IV and MIMIC-CXR [1,3]. The *subject\_id* value provides an implicit link between the datasets; that is all three databases refer to the same individual with the same *subject\_id*. All ED stays in MIMIC-IV-ED, represented by *stay\_id*, are present in the MIMIC-IV *transfers* table. Linking to MIMIC-IV, for example, would provide approximate age for ED patients as this data is available in the *patients* table present in MIMIC-IV. Laboratory measurements for ED patients would be available in the *labevents* table of the hosp module in MIMIC-IV, prescribed medications would be available in the *prescriptions* table of the hosp module in MIMIC-IV, and so on. Emergency department patients who are eventually admitted to the intensive care unit would have information regarding their subsequent ICU stay in the icu module of MIMIC-IV. As a result, MIMIC-IV-ED may be used to acquire pre-ICU information for critically ill patients in MIMIC-IV. MIMIC-IV covers a wider time frame than MIMIC-IV-ED, and as such not all emergency department stays in MIMIC-IV will be present in MIMIC-IV-ED, but all ED admissions in MIMIC-IV-ED will be present in MIMIC-IV.

Patients within MIMIC-CXR are a subset of patients within MIMIC-IV-ED. As a result, many ED patients who have a chest x-ray ordered will have the image and radiology report available in MIMIC-CXR. Note that not all ED patients will have x-rays in MIMIC-CXR as MIMIC-IV-ED covers a larger time frame, but almost all ED stays which have x-rays in MIMIC-CXR will have the associated stay present in MIMIC-IV-ED.

## Limitations

Data contained within MIMIC-IV-ED are collected during routine clinical care, and their use for research is secondary to their use in clinical care. The data may contain implicit biases as a result of local data collection practices, implausible values for measurements, and missing documentation for provided treatments. Many interventions, including major events such as endotracheal intubation, are not documented clearly. Researchers should take care to address these issues in their work.

---

## Release Notes

### MIMIC-IV-ED v2.2

MIMIC-IV-ED v2.2 was released on January 5, 2023. It removed a subset of *subject\_id* which will be retained internally as a test set. Future data releases will exclude these patients.

- *edstays* - Removed 22,625 *stay\_id* from the table.
- Other tables will have rows removed to reflect the removal of the aforementioned *stay\_id*. Final row counts are available in the validation scripts published with the MIMIC Code Repository [8].

# MIMIC-IV-ED v2.0

MIMIC-IV-ED v2.0 was released in May, 2022. This was an improvement to MIMIC-IV-ED with additional datatypes. As there were schema changes, the major version was incremented.

Changes include:

- Added additional columns to the *edstays* table.
  - gender
  - race
  - arrival\_transport
  - disposition
- Fixed a bug where the outtime for stays with no subsequent hospitalization was incorrect. This resulted in all *edstays* rows with a NULL *hadm\_id* having an apparent ED stay of minutes or less. The outtime column has been corrected.
- The *pain* column of the *triage* table is now free-text, and now includes free-text entries which are not valid numbers. This is more consistent with the *pain* column in the *vitalsign* table, which was already free-text.

## MIMIC-IV-ED v1.0

MIMIC-IV-ED v1.0 was released on June 3rd, 2021. The initial release of MIMIC-IV-ED contained six tables: *edstays*, *diagnosis*, *medrecon*, *pyxis*, *triage*, and *vitalsign*.

---

## Ethics

The collection of patient information and creation of the research resource was reviewed by the Institutional Review Board at the Beth Israel Deaconess Medical Center, who granted a waiver of informed consent and approved the data sharing initiative (IRB #2001P001699).

---

## Acknowledgements

We would like to thank the Beth Israel Deaconess Medical Center for their continued collaboration and support of MIMIC. In particular we thank Carolyn Conti, Alvin Gayles, Ayad Shammout, and Lu Shen for their help with data extraction.

---

## Conflicts of Interest

Nothing to declare.

---

## References

1. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2021). MIMIC-IV (version 1.0). PhysioNet. <https://doi.org/10.13026/s6n6-xd98>.
2. Health Insurance Portability and Accountability Act [HIPAA] of 1996, Pub. L. No. 104-191. <https://www.congress.gov/104/plaws/publ191/PLAW-104publ191.pdf>
3. Johnson, A., Pollard, T., Mark, R., Berkowitz, S., & Horng, S. (2019). MIMIC-CXR Database (version 2.0.0). PhysioNet. <https://doi.org/10.13026/C2JT1Q>.
4. Johnson, A., Lungren, M., Peng, Y., Lu, Z., Mark, R., Berkowitz, S., & Horng, S. (2019). MIMIC-CXR-JPG - chest radiographs with structured labels (version 2.0.0). PhysioNet. <https://doi.org/10.13026/8360-t248>.
5. Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J. et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data* 6, 317 (2019). <https://doi.org/10.1038/s41597-019-0322-0>
6. Johnson AEW, Bulgarelli L, and Pollard T. 2020. Deidentification of free-text medical records using pre-trained bidirectional transformers. In Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL '20). Association for Computing Machinery, New York, NY, USA, 214–221. DOI:<https://doi.org/10.1145/3368555.3384455>
7. Pyxis Medstation Website. <https://www.bd.com/en-us/offerings/capabilities/medication-and-supply-management/medication-and->

[supply-management-technologies/pyxis-medication-technologies/pyxis-medstation-es-system](#) [Accessed: 10 April 2021]

8. MIMIC Code Repository on GitHub. <https://github.com/MIT-LCP/mimic-code/> [Accessed: 1 May 2021]

9. Alistair E W Johnson, David J Stone, Leo A Celi, Tom J Pollard, The MIMIC Code Repository: enabling reproducibility in critical care research, Journal of the American Medical Informatics Association, Volume 25, Issue 1, January 2018, Pages 32–39, <https://doi.org/10.1093/jamia/ocx084>

---

## Files

This is a restricted-access resource. To access the files, you must fulfill all of the following requirements:

- be a [credentialed user](#)
- complete required training:
  - [CITI Data or Specimens Only Research](#)  
You may submit your training [here](#).
- [sign the data use agreement](#) for the project

PhysioNet is a repository of freely-available medical research data, managed by the MIT Laboratory for Computational Physiology.

Supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH grant number R01EB030362.

For more accessibility options, see the [MIT Accessibility Page](#).

[Back to top](#)