Database | Credentialed Access

MIMIC-III Clinical Database

Alistair Johnson ⓘ , **Tom Pollard** ⓘ , **Roger Mark** ⓘ

MIMIC-III available on the cloud *(Sept. 30, 2019, 2:28 p.m.)*

The MIMIC-III database is now available on two major cloud platforms: Google Cloud Platform (GCP) and Amazon Web Services (AWS). To access the data on the cloud, simply add the relevant cloud identifier to your PhysioNet profile. Further instructions are available on the MIMIC-III website.

Tutorials are available for using MIMIC-III with BigQuery (GCP) and using MIMIC-III with Athena (AWS).

When using this resource, please cite: (show more options)
Johnson, A., Pollard, T., & Mark, R. (2016). MIMIC-III Clinical Database (version 1.4). *PhysioNet*. RRID:SCR_007345. https://doi.org/10.13026/C2XW26

Additionally, please cite the original publication:
Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.

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. RRID:SCR_007345.

Abstract

MIMIC-III is a large, freely-available database comprising deidentified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. The database includes information such as demographics, vital sign measurements made at the bedside (~1 data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (including post-hospital discharge).

MIMIC supports a diverse range of analytic studies spanning epidemiology, clinical decision-rule improvement, and electronic tool development. It is notable for three factors: it is freely available to researchers worldwide; it encompasses a diverse and very large population of ICU patients; and it contains highly granular data, including vital signs, laboratory results, and medications.

Background

In recent years there has been a concerted move towards the adoption of digital health record systems in hospitals. In the US, for example, the number of non-federal acute care hospitals with basic digital systems increased from 9.4 to 75.5% over the 7 year period between 2008 and 2014 [1].

Despite this advance, interoperability of digital systems remains an open issue, leading to challenges in data integration. As a result, the potential that hospital data offers in terms of understanding and improving care is yet to be fully realized. In parallel, the scientific research community is increasingly coming under criticism for the lack of reproducibility of studies [2].

Methods

MIMIC-III integrates deidentified, comprehensive clinical data of patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts, and makes it widely accessible to researchers internationally under a data use agreement. The open nature of the data allows clinical studies to be reproduced and improved in ways that would not otherwise be possible.

The MIMIC-III database was populated with data that had been acquired during routine hospital care, so there was no associated burden on caregivers and no interference with their workflow. Data was downloaded from several sources, including:

- archives from critical care information systems.
- hospital electronic health record databases.
- Social Security Administration Death Master File.

Two different critical care information systems were in place over the data collection period: Philips CareVue Clinical Information System (models M2331A and M1215A; Philips Health-care, Andover, MA) and iMDsoft MetaVision ICU (iMDsoft, Needham, MA). These systems were the source of clinical data such as:

- time-stamped nurse-verified physiological measurements (for example, hourly documentation of heart rate, arterial blood pressure, or respiratory rate);
- documented progress notes by care providers;
- continuous intravenous drip medications and fluid balances.

With exception to data relating to fluid intake, which differed significantly in structure between the CareVue and MetaVision systems, data was merged when building the database tables. Data which could not be merged is given a suffix to denote the data source. For example, inputs for patients monitored with the CareVue system are stored in INPUTEVENTS_CV, whereas inputs for patients monitored with the Metavision system are stored in INPUTEVENTS_MV. Additional information was collected from hospital and laboratory health record systems, including:

- patient demographics and in-hospital mortality.
- laboratory test results (for example, hematology, chemistry, and microbiology results).
- discharge summaries and reports of electrocardiogram and imaging studies.
- billing-related information such as International Classification of Disease, 9th Edition (ICD-9) codes, Diagnosis Related Group (DRG) codes, and Current Procedural Terminology (CPT) codes.

Out-of-hospital mortality dates were obtained using the Social Security Administration Death Master File.

Before data was incorporated into the MIMIC-III database, it was first deidentified in accordance with Health Insurance Portability and Accountability Act (HIPAA) standards using structured data cleansing and date shifting. The deidentification process for structured data required the removal of all eighteen of the identifying data elements listed in HIPAA, including fields such as patient name, telephone number, address, and dates. In particular, dates were shifted into the future by a random offset for each individual patient in a consistent manner to preserve intervals, resulting in stays which occur sometime between the years 2100 and 2200. Time of day, day of the week, and approximate seasonality were conserved during date shifting. Dates of birth for patients aged over 89 were shifted to obscure their true age and comply with HIPAA regulations: these patients appear in the database with ages of over 300 years.

Protected health information was removed from free text fields, such as diagnostic reports and physician notes, using a rigorously evaluated deidentification system based on extensive dictionary look-ups and pattern-matching with regular expressions. The components of this deidentification system are continually expanded as new data is acquired.

The project was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). Requirement for individual patient consent was waived because the project did not impact clinical care and all protected health information was deidentified.

Data Description

MIMIC-III is a relational database consisting of 26 tables. 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 ICUSTAY_ID refers to a unique admission to an intensive care unit.

Charted events such as notes, laboratory tests, and fluid balance are stored in a series of 'events' tables. For example the OUTPUTEVENTS table contains all measurements related to output for a given patient, while the LABEVENTS table contains laboratory test results for a patient.

Tables prefixed with 'D_' are dictionary tables and provide definitions for identifiers. For example, every row of CHARTEVENTS is associated with a single ITEMID which represents the concept measured, but it does not contain the actual name of the measurement. By joining CHARTEVENTS and D_ITEMS on ITEMID, it is possible to identify the concept represented by a given ITEMID.

Developing the MIMIC data model involved balancing simplicity of interpretation against closeness to ground truth. As such, the model is a reflection of underlying data sources, modified over iterations of the MIMIC database in response to user feedback. Care has been taken to avoid making assumptions about the underlying data when carrying out transformations, so MIMIC-III closely represents the raw hospital data.

Broadly speaking, five tables are used to define and track patient stays: ADMISSIONS; PATIENTS; ICUSTAYS; SERVICES; and TRANSFERS. Another five tables are dictionaries for cross-referencing codes against their respective definitions: D_CPT; D_ICD_DIAGNOSES; D_ICD_PROCEDURES; D_ITEMS; and D_LABITEMS. The remaining tables contain data associated with patient care, such as physiological measurements, caregiver observations, and billing information.

In some cases it would be possible to merge tables—for example, the D_ICD_PROCEDURES and CPTEVENTS tables both contain detail relating to procedures and could be combined—but our approach is to keep the tables independent for clarity, since the data sources are significantly different. Rather than combining the tables within MIMIC data model, we suggest researchers develop database views and transforms as appropriate.

Usage Notes

MIMIC-III is provided as a collection of comma separated value (CSV) files, along with scripts to help with importing the data into database systems including PostreSQL, MySQL, and MonetDB. As the database contains detailed information regarding the clinical care of patients, it must be treated with appropriate care and respect. Researchers are required to formally request access via a process documented on the [MIMIC website](). There are two key steps that must be completed before access is granted:

- the researcher must complete a recognized course in protecting human research participants that includes Health Insurance Portability and Accountability Act (HIPAA) requirements.

- the researcher must sign a data use agreement, which outlines appropriate data usage and security standards, and forbids efforts to identify individual patients.

Approval requires at least a week. Once an application has been approved the researcher will receive emails containing instructions for downloading the database from PhysioNetWorks, a restricted access component of PhysioNet.

Release Notes
MIMIC-III v1.4

The current version of the database is v1.4. When referencing this version, we recommend using the full title: MIMIC-III v1.4.

MIMIC-III v1.4 was released on 2 September 2016. It was a major release enhancing data quality and providing a large amount of additional data for Metavision patients.

Issues addressed include:

- #88 - Text data sourced from drop down menus or "pick lists" has been added to the database for Metavision patients. All data was added to the chartevents table, increasing the raw (uncompressed) size of the datafile by 5 GB.Users interested in the exact concepts added can query them as follows: select * from d_items where linksto = 'chartevents' and dbsource = 'metavision' and param_type = 'Text'
- #201 - Data in tables may now only occur within 1 year of a patient's hospital admission admittime and dischtime. Observations outside this range corresponded to typographical errors in the charttime, and this unreliable data has been removed.
- #197 - A small issue where a single observation was duplicated and assigned to two icustay_id (and sometimes two hadm_id) has been corrected: this affected 37 icustay_id and 80 hadm_id. Thanks @matteobonvini on GitHub for the report.
- #191 - A subset of patients were missing data due to the transition from CareVue to Metavision: this data has been restored.
- #193 - Some laboratory concepts (particularly CD counts) were conflated (i.e. percentages mixed with counts) - these have been corrected. Thanks Yuqi Si for the report (via StackExchange).
- #196 - The amount column in inputevents_mv did not correspond to amountuom (unit of measurement) - instead it corresponded to a "base" unit (e.g. grams instead of milligrams). Data in the amount has been updated to match the amountuom where appropriate. This also effected the procedureevents_mv table. Thanks @ngageorange on GitHub for the report.
- #198 - NULL values for character fields are now properly represented in the CSV file as empty fields with no double quotes. Database systems such as PostgreSQL will now recognize these fields as null when copying with the NULL " qualifier.
- #199 - Addendums to discharge summaries in the notes can now be distinguished in the noteevents table: their category has remained 'Discharge summary', but their description has been changed from 'Report' to 'Addendum'
- #194 - The LOINC code for Total CO_2 has been corrected (was 1959-6 total bicarb, is now 34728-6 total CO2)
- #186 - Data has been harmonized into EST.
- #180 - Some patients were erroneously labelled as in care unit TSICU: these have been corrected to CSRU
- #189 - Schema change: The has_ioevents_data column in admissions has been removed as it was no longer meaningful.
- #185 - Three text concepts (GCS, code status, RASS) had a null linksto field in d_items, even though the data existed in chartevents. This has been corrected. Note these were the only three text concepts available for Metavision patients in MIMIC-III v1.3 (see #88).

- #60 – The date shift for dates in the text field of noteevents has been corrected: before it corresponded to a different date shifted used in MIMIC-II (easily identifiable by years occurring between 2500-3500). It now matches the dates in the structured data.
- #188 – Negative microbiology cultures are now recorded in micriobiologyevents – negative cultures can be identified by a NULL value for the organism name (org_name).
- Be sure to validate the checksum of the resultant file to ensure you have the correct version.

MIMIC-III v1.3

MIMIC-III v1.3 was released on 10 December 2015. It was a minor release enhancing the consistency of the dataset.

Issues addressed include:

- #175 – A new value for DRG_VERSION was added to the DRGCODES table to clarify why the same code matched to multiple descriptions.
- #174 – The EDTIMEOUT column was renamed to EDOUTTIME in the ADMISSIONS table for consistency with other timestamp columns.
- #173 – The UOM column was renamed to VALUEUOM in the CHARTEVENTS, DATETIMEEVENTS, and LABEVENTS tables for consistency with other UOM columns.
- #172, #177 – Several careunit acronyms were merged in the TRANSFERS and ICUSTAYS tables for ease of interpretation.
- #168 – A set of ITEMIDs in the INPUTEVENTS_CV table were inappropriately low (<30000), they have no been corrected.
- #167 – Duplicate radiology reports were removed from the NOTEEVENTS table. These duplicates were present in the raw data.
- #166 – The DBSOURCE column was corrected from Metavision to CareVue for a set of patients in the TRANSFERS and ICUSTAYS tables.

MIMIC-III v1.2

MIMIC-III v1.2 was released on 10 November 2015. MIMIC-III v1.2 was a major release providing both bug fixes and a large amount of additional data.

Major issues addressed, including additional data made available:

- #130 and #135 – Duplicate data in various events tables with CGID has been removed
- #132 – Hospital expire flag was in the wrong table – now moved to ADMISSIONS table. EXPIRE_FLAG added to PATIENTS table.
- #137 – The ITEMID for input/output items has been properly shifted to range between 30000-40000. Previously it incorrectly ranged between 1-5000, and as a result did not match the dictionary entries in D_ITEMS.
- #141 – CHARTTIME and STORETIME have been added to NOTEEVENTS, if available. No times were available for ECG and echo reports (only the date).
- #144 – Dates of birth for patients > 89 are now shifted by 300 years for clarity.
- #151 – The time of emergency department registration and exit has been added to the admissions table, where available.
- #154 – A new table, PROCEDUREEVENTS_MV, has been added. This table contains information regarding the start and stop time for various procedures for Metavision patients. Procedures include x-rays, ventilation, dialysis, and others.
- #158 – The IOEVENTS table has been split into three tables: INPUTEVENTS_CV (CareVue patients only), INPUTEVENTS_MV (Metavision patients only), and OUTPUTEVENTS (all patients).
- #162 – 5,795,842 rows of data corresponding to yes/no answers have been added for Metavision patients
- #164 – 10,140 rows of "Non Iv meds" (e.g. vancomycin) have been added for Metavision patients

Minor issues addressed:

- #126 – ROW_ID, CGID, ORDERID and LINKORDERID now stored as INT instead of BIGINT
- #134 – The CR/LF characters which prefixed notes in NOTEEVENTS have been removed
- #136 – The VOLUME column in the events tables for inputs has been changed to AMOUNT
- #139 – The units for certain solution volumes, erroneously recorded as rate units, have been corrected
- #140 – The ORIGINALAMOUNT column no longer exists for rows with drug rates, only with drug volumes
- #148 – Removed RECORD_ID from NOTEEVENTS as it was redundant
- #149 – ICUSTAYEVENTS has been renamed ICUSTAYS
- #152 – The STARTTIME and ENDTIME columns have been renamed STARTDATE and ENDDATE for the PRESCRIPTIONS table to reflect the lack of time information.
- #155 – LINKSTO column in D_ITEMS has been corrected: now correctly refers to PROCEDUREEVENTS_MV, INPUTEVENTS_MV, INPUTEVENTS_CV, and OUTPUTEVENTS.
- #156 – The SEQUENCE column in DIAGNOSES_ICD and PROCEDURES_ICD has been renamed SEQ_NUM.
- #159 – A CONCEPTID column has been added to D_ITEMS for future ontology mapping/data harmonization.
- #163 – The CODE column has been removed from MICROBIOLOGYEVENTS and D_ITEMS as it was redundant to ITEMID

MIMIC-III v1.1

MIMIC-III v1.1 was released on 24 September 2015. It was primarily a bug fix release, and addresses the following issues:

- #116 – CGID was incorrect in the DATETIMEEVENTS, CHARTEVENTS, IOEVENTS and NOTEEVENTS tables. It has now been corrected.
- #117 – VALUENUM for GCS verbal response measurements has been corrected for Metavision (it was offset by -1).
- #118 – VALUENUM for all GCS measurements in CareVue is no longer null, and contains the appropriate value in the scale.
- #120 – DOD was incorrectly set to DOB – this has been fixed.
- #121 – IOEVENTS contained incorrect units for certain drugs (sometimes the unit was a rate when the actual observation was an amount, e.g. listed as "mcgkgmin" when it should have been "mg").
- #122 – DBSOURCE in the TRANSFERS and ICUSTAYEVENTS tables has been corrected – originally it only contained 'metavision' when a patient was in the ICU, so the same patient would be listed as 'carevue' when out of the ICU and 'metavision' when inside the ICU.
- #123 – Precision in the IOEVENTS table has been fixed at 10 decimal places.

MIMIC-III v1.0

MIMIC-III v1.0 was released on 25 August 2015. It was a preliminary version and not widely publicized to allow for internal testing. As this was the first release of the database, and the successor of the MIMIC-II database, no changes are listed here. An overview of changes between MIMIC-II and MIMIC-III is provided here.

Acknowledgements

Conflicts of Interest

The authors declare no competing financial interests.

References

1. Charles, D., King, J., Patel, V. & Furukawa, M. Adoption of Electronic Health record Systems among U.S. Non-federal Acute Care Hospitals. ONC Data Brief No. 9, 1–9 (2013).
2. Collins, F. S. & Tabak, L. A. NIH plans to enhance reproducibility. Nature 505, 612–613 (2014).

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 1.4):
https://doi.org/10.13026/C2XW26

DOI (latest version):
https://doi.org/10.13026/cd7z-wg25

Topics:
clinical   intensive care   critical care   natural language processing   machine learning

Project Website:
↗ https://mimic.physionet.org

Corresponding Author

You must be logged in to view the contact information.

Versions

1.4	Sept. 4, 2016

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

Navigation

Discover Data

Share Data

About

News

Explore

Data

Software

Challenges

Tutorials

Accessibility