MIMIC III & IV

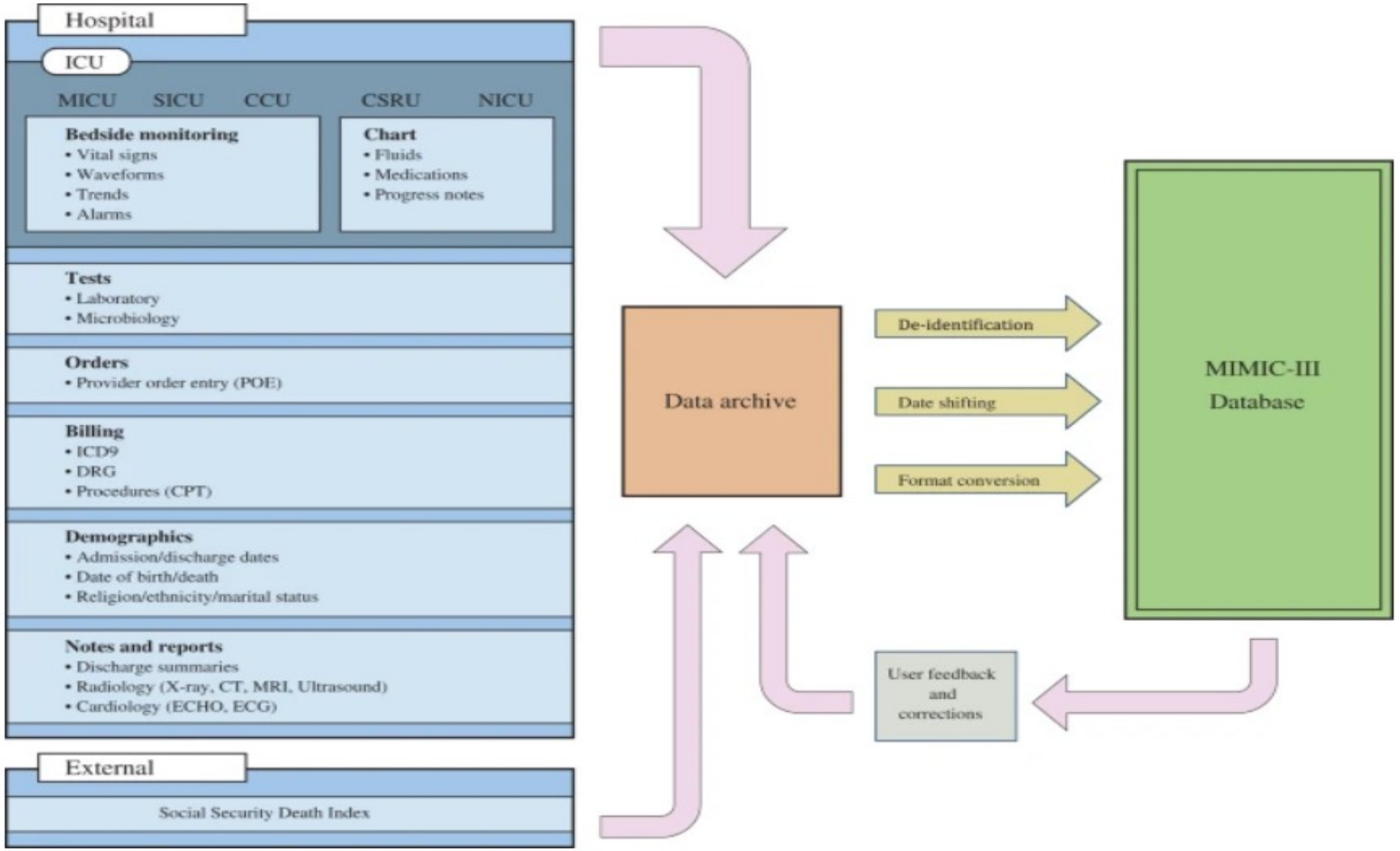
AI in Health

UT Austin

MIMIC III ICU Data

- MIMIC (Medical Information Mart for Intensive Care) is an openly available dataset developed by the MIT Lab for Computational Physiology, comprising deidentified health data associated with ~60,000 intensive care unit admissions. It includes demographics, vital signs, laboratory tests, medications, and more.
- The latest version of MIMIC is MIMIC-III v1.4, which comprises 61,532 intensive care unit stays at Beth Israel Deaconess Medical Center in Boston, Massachusetts: 53,432 stays for adult patients and 8,100 for neonatal patients. The data spans June 2001 - October 2012.

Overview of MIMIC III ICU data



Data statistics

Critical care unit	CCU	CSRU	MICU	SICU	TSICU	Total
Distinct patients, no. (% of total admissions)	5,674 (14.7%)	8,091 (20.9%)	13,649 (35.4%)	6,372 (16.5%)	4,811 (12.5%)	38,597 (100%)
Hospital admissions, no. (% of total admissions)	7,258 (14.6%)	9,156 (18.4%)	19,770 (39.7%)	8,110 (16.3%)	5,491 (11.0%)	49,785 (100%)
Distinct ICU stays, no. (% of total admissions)	7,726 (14.5%)	9,854 (18.4%)	21,087 (39.5%)	8,891 (16.6%)	5,865 (11.0%)	53,423 (100%)
Age, years, median (Q1-Q3)	70.1 (58.4–80.5)	67.6 (57.6–76.7)	64.9 (51.7–78.2)	63.6 (51.4–76.5)	59.9 (42.9–75.7)	65.8 (52.8–77.8)
Gender, male, % of unit stays	4,203 (57.9%)	6,000 (65.5%)	10,193 (51.6%)	4,251 (52.4%)	3,336 (60.7%)	27,983 (55.9%)
ICU length of stay, median days (Q1-Q3)	2.2 (1.2–4.1)	2.2 (1.2–4.0)	2.1 (1.2–4.1)	2.3 (1.3–4.9)	2.1 (1.2–4.6)	2.1 (1.2–4.6)
Hospital length of stay, median days (Q1-Q3)	5.8 (3.1–10.0)	7.4 (5.2–11.4)	6.4 (3.7–11.7)	7.9 (4.4–14.2)	7.4 (4.1–13.6)	6.9 (4.1–11.9)
ICU mortality, percent of unit stays	685 (8.9%)	353 (3.6%)	2,222 (10.5%)	813 (9.1%)	492 (8.4%)	4,565 (8.5%)
Hospital mortality, percent of unit stays	817 (11.3%)	424 (4.6%)	2,859 (14.5%)	1,020 (12.6%)	628 (11.4%)	5,748 (11.5%)

CCU is Coronary Care Unit; CSRU is Cardiac Surgery Recovery Unit; MICU is Medical Intensive Care Unit; SICU is Surgical Intensive Care Unit; TSICU is Trauma Surgical Intensive Care Unit.

ICD9 code coverage by patients 16 years older

Critical care unit	CCU stays, No. (% by unit)	CSRU stays, No. (% by unit)	MICU stays, No. (% by unit)	SICU stays, No. (% by unit)	TSICU stays, No. (% by unit)	Total stays, No. (% by unit)
Infectious and parasitic diseases, i.e., septicemia, other infectious and parasitic diseases, etc., (001–139)	305 (4.2%)	72 (0.8%)	3,229 (16.7%)	448 (5.6%)	152 (2.8%)	4,206 (8.6%)
Neoplasms of digestive organs and intrathoracic organs, etc., (140–239)	126 (1.8%)	287 (3.2%)	1,415 (7.3%)	1,225 (15.3%)	466 (8.6%)	3,519 (7.2%)
Endocrine, nutritional, metabolic, and immunity (240–279)	104 (1.4%)	36 (0.4%)	985 (5.1%)	178 (2.2%)	54 (1.0%)	1,357 (2.8%)
Diseases of the circulatory system, i.e., ischemic heart diseases, diseases of pulmonary circulation, dysrhythmias, heart failure, cerebrovascular diseases, etc., (390–459)	5,131 (71.4%)	7,138 (78.6%)	2,638 (13.6%)	2,356 (29.5%)	684 (12.6%)	17,947 (36.6%)
Pulmonary diseases, i.e., pneumonia and influenza, chronic obstructive pulmonary disease, etc., (460–519)	416 (5.8%)	141 (1.6%)	3,393 (17.5%)	390 (4.9%)	225 (4.1%)	4,565 (9.3%)
Diseases of the digestive system (520–579)	264 (3.7%)	157 (1.7%)	3,046 (15.7%)	1,193 (14.9%)	440 (8.1%)	5,100 (10.4%)
Diseases of the genitourinary system, i.e., nephritis, nephrotic syndrome, nephrosis, and other diseases of the genitourinary system (580–629)	130 (1.8%)	14 (0.2%)	738 (3.8%)	101 (1.3%)	31 (0.6%)	1,014 (2.1%)
Trauma (800–959)	97 (1.3%)	494 (5.4%)	480 (2.5%)	836 (10.5%)	2,809 (51.7%)	4,716 (9.6%)
Poisoning by drugs and biological substances (960–979)	50 (0.7%)	2 (0.0%)	584 (3.0%)	58 (0.7%)	11 (0.2%)	705 (1.4%)
Other	565 (7.9%)	739 (8.1%)	2,883 (14.9%)	1,204 (15.1%)	563 (10.4%)	5,954 (12.1%)
Total	7,188 (14.6%)	9,080 (18.5%)	19,391 (39.5%)	7,989 (16.3%)	5,435 (11.1%)	49,083 (100%)

Kinds of data available at MIMIC III

Class of data	Description
Billing	Coded data recorded primarily for billing and administrative purposes. Includes Current Procedural Terminology (CPT) codes, Diagnosis-Related Group (DRG) codes, and International Classification of Diseases (ICD) codes.
Descriptive	Demographic detail, admission and discharge times, and dates of death.
Dictionary	Look-up tables for cross referencing concept identifiers (for example, International Classification of Diseases (ICD) codes) with associated labels.
Interventions	Procedures such as dialysis, imaging studies, and placement of lines.
Laboratory	Blood chemistry, hematology, urine analysis, and microbiology test results.
Medications	Administration records of intravenous medications and medication orders.
Notes	Free text notes such as provider progress notes and hospital discharge summaries.
Physiologic	Nurse-verified vital signs, approximately hourly (e.g., heart rate, blood pressure, respiratory rate).
Reports	Free text reports of electrocardiogram and imaging studies.

MIMIC data source

- Data was downloaded from several sources, including:
  - archives from critical care information systems.
    - 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.
  - hospital electronic health record databases.
    - 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.
  - Social Security Administration Death Master File.
    - Out-of-hospital mortality dates

Related Projects

- Several projects are ongoing to map concepts within the MIMIC database to standardized dictionaries:
  - Researchers at the National Library of Medicine National Institutes of Health have mapped laboratory tests and medications in MIMIC-II to LOINC and RxNorm, respectively.
  - Efforts are also underway to transform MIMIC to common data models, such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model, to support the application of standardized tools and methods.

MIMIC Data

- 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.

Each table details

- https://mimic.mit.edu/docs/iii/tables/admissions/

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ADMITTIME	TIMESTAMP(0)
DISCHTIME	TIMESTAMP(0)
DEATHTIME	TIMESTAMP(0)
ADMISSION_TYPE	VARCHAR(50)
ADMISSION_LOCATION	VARCHAR(50)
DISCHARGE_LOCATION	VARCHAR(50)

SUBJECT_ID, HADM_ID

Each row of this table contains a unique HADM_ID, which represents a single patient's admission to the hospital. HADM_ID ranges from 1000000 - 1999999. It is possible for this table to have duplicate SUBJECT_ID, indicating that a single patient had multiple admissions to the hospital. The ADMISSIONS table can be linked to the PATIENTS table using SUBJECT_ID.

MIMC III Tables

Table name	Description
ADMISSIONS	Every unique hospitalization for each patient in the database (defines HADM_ID).
CALLOUT	Information regarding when a patient was cleared for ICU discharge and when the patient was actually discharged.
CAREGIVERS	Every caregiver who has recorded data in the database (defines CGID).
CHARTEVENTS	All charted observations for patients.
CPTEVENTS	Procedures recorded as Current Procedural Terminology (CPT) codes.
D_CPT	High level dictionary of Current Procedural Terminology (CPT) codes.
D_ICD_DIAGNOSES	Dictionary of International Statistical Classification of Diseases and Related Health Problems (ICD-9) codes relating to diagnoses.
D_ICD_PROCEDURES	Dictionary of International Statistical Classification of Diseases and Related Health Problems (ICD-9) codes relating to procedures.
D_ITEMS	Dictionary of local codes ('ITEMIDs') appearing in the MIMIC database, except those that relate to laboratory tests.
D_LABITEMS	Dictionary of local codes ('ITEMIDs') appearing in the MIMIC database that relate to laboratory tests.
DATETIMEEVENTS	All recorded observations which are dates, for example time of dialysis or insertion of lines.
DIAGNOSES_ICD	Hospital assigned diagnoses, coded using the International Statistical Classification of Diseases and Related Health Problems (ICD) system.
DRGCODES	Diagnosis Related Groups (DRG), which are used by the hospital for billing purposes.

MIMC III Tables

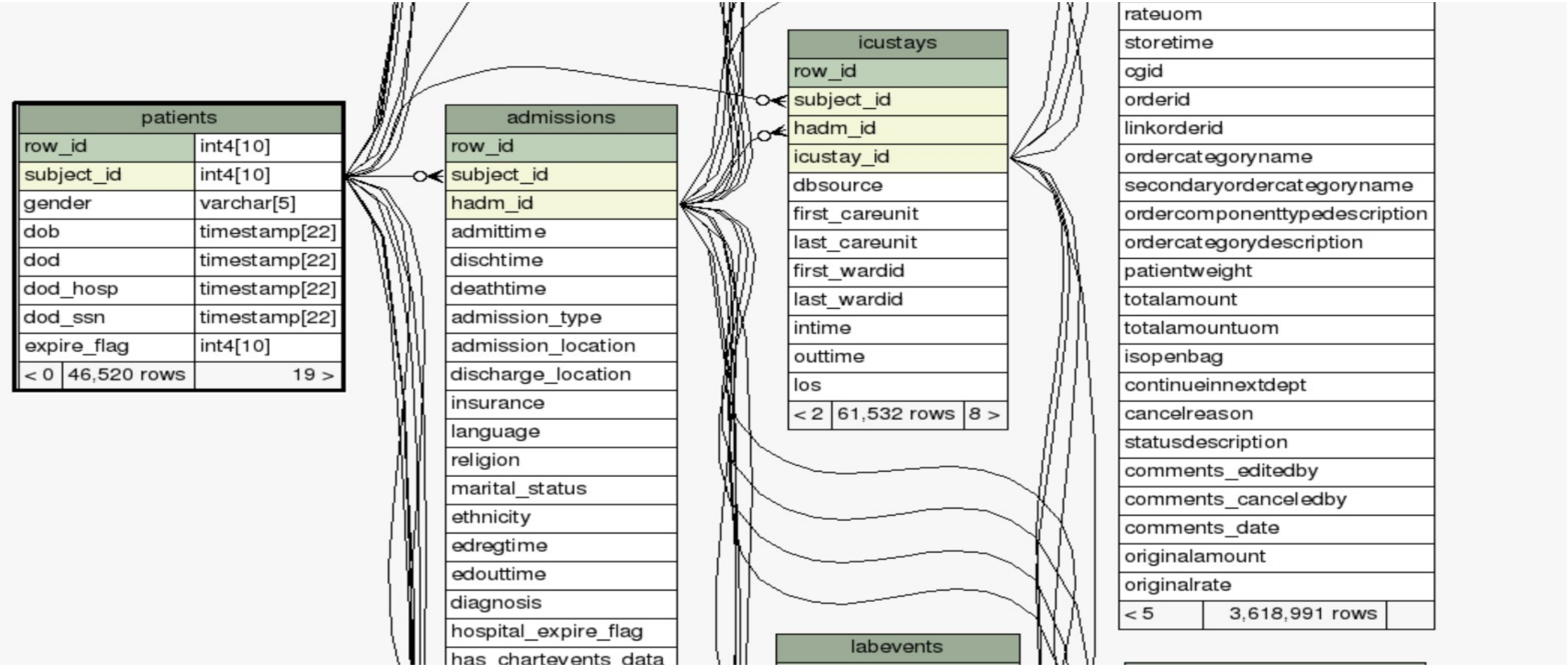
ICUSTAYS	Every unique ICU stay in the database (defines ICUSTAY_ID).
INPUTEVENTS_CV	Intake for patients monitored using the Philips CareVue system while in the ICU, e.g., intravenous medications, enteral feeding, etc.
INPUTEVENTS_MV	Intake for patients monitored using the iMDSoft MetaVision system while in the ICU, e.g., intravenous medications, enteral feeding, etc.
OUTPUTEVENTS	Output information for patients while in the ICU.
LABEVENTS	Laboratory measurements for patients both within the hospital and in outpatient clinics.
MICROBIOLOGYEVENTS	Microbiology culture results and antibiotic sensitivities from the hospital database.
NOTEEVENTS	Deidentified notes, including nursing and physician notes, ECG reports, radiology reports, and discharge summaries.
PATIENTS	Every unique patient in the database (defines SUBJECT_ID).
PRESCRIPTIONS	Medications ordered for a given patient.
PROCEDUREEVENTS_MV	Patient procedures for the subset of patients who were monitored in the ICU using the iMDSoft MetaVision system.
PROCEDURES_ICD	Patient procedures, coded using the International Statistical Classification of Diseases and Related Health Problems (ICD) system.
SERVICES	The clinical service under which a patient is registered.
TRANSFERS	Patient movement from bed to bed within the hospital, including ICU admission and discharge.

Common Tables

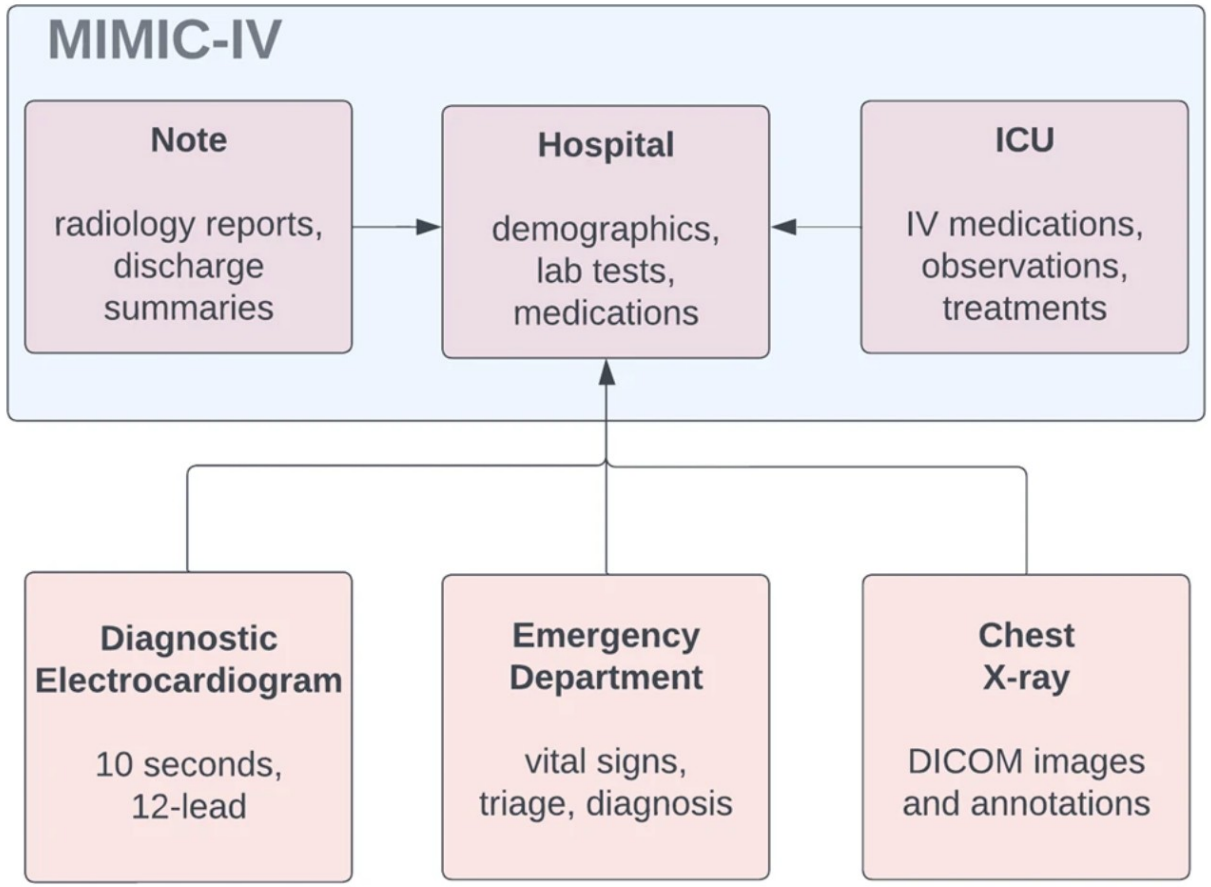
- Define and track patient stays:
  - ADMISSIONS; PATIENTS; ICUSTAYS; SERVICES; and TRANSFERS.
- Cross-referencing codes against their respective definitions:
  - D_CPT; D_ICD_DIAGNOSES; D_ICD_PROCEDURES; D_ITEMS; and D_LABITEMS.
- The remaining tables contain data associated with patient care, such as physiological measurements, caregiver observations, and billing information.

MIMIC III Data Schema

- Data Schema: https://mit-lcp.github.io/mimic-schema-spy/



MIMIC IV



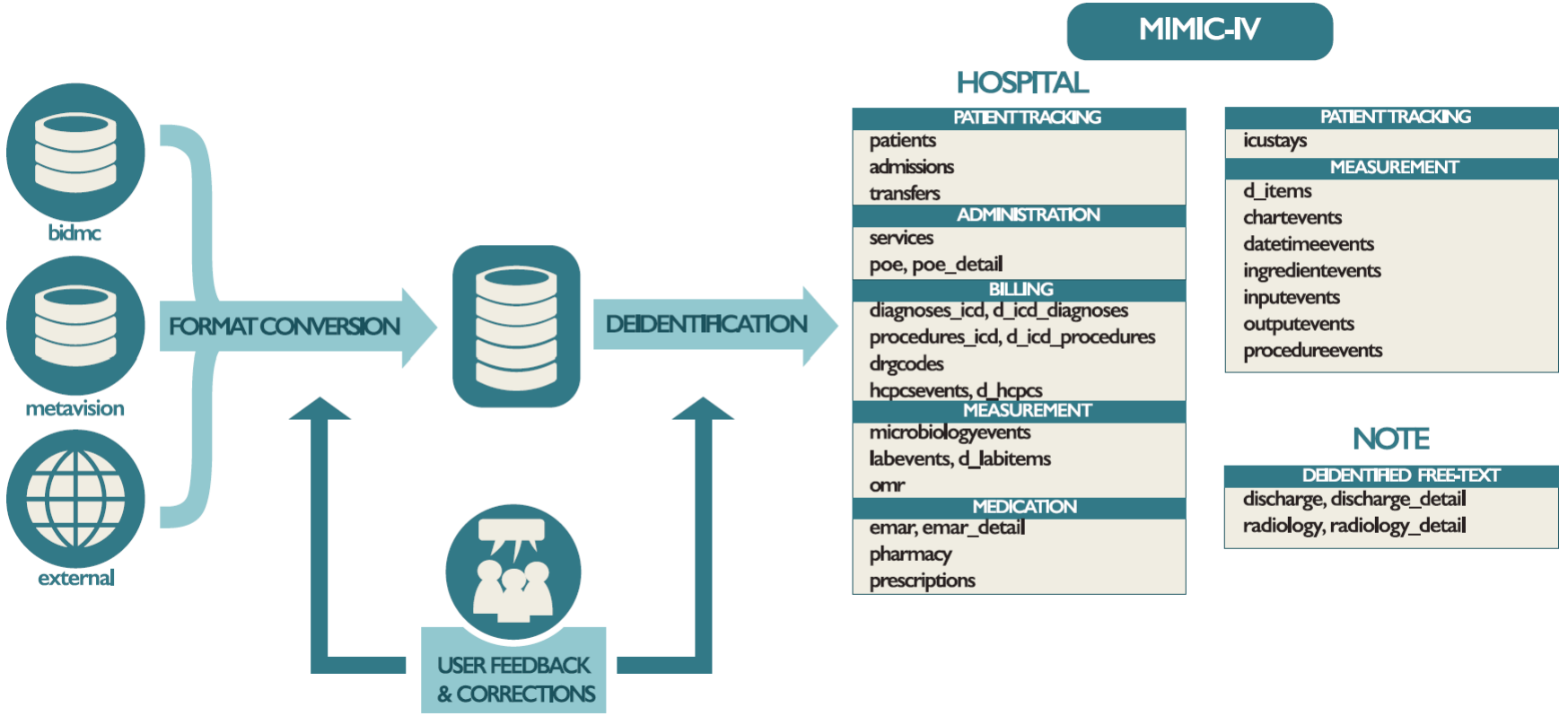
MIMIC-IV

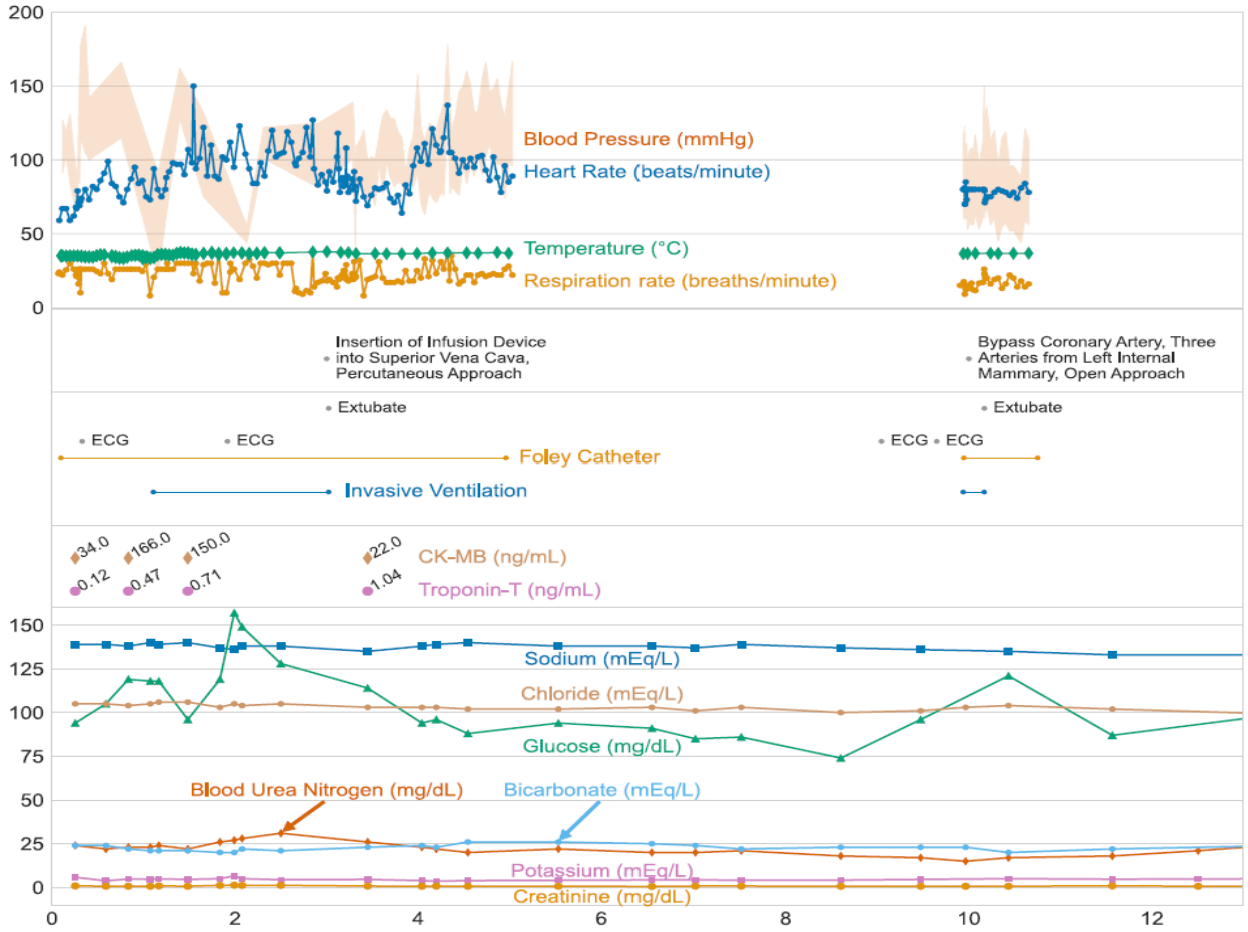
- 2008–2019
- ICD 9 and ICD 10
- three modules: **hosp**, **icu**, and **note**

	Hospital admissions	ICU admissions
Number of stays	431,231	73,181
Unique patients	180,733	50,920
Age, mean (SD)	58.8 (19.2)	64.7 (16.9)
Female Administrative Gender, n (%)	224,990 (52.2)	32,363 (44.2)
Insurance, n (%)		
Medicaid	41,330 (9.6)	5,528 (7.6)
Medicare	160,560 (37.2)	33,091 (45.2)
Other	229,341 (53.2)	34,562 (47.2)
Hospital length of stay, mean (SD)	4.5 (6.6)	11.0 (13.3)
In-hospital mortality, n (%)	8,974 (2.1)	8,519 (11.6)
One year mortality, n (%)	106,218 (24.6)	28,274 (38.6)

https://www.nature.com/articles/s41597-022-01899-x

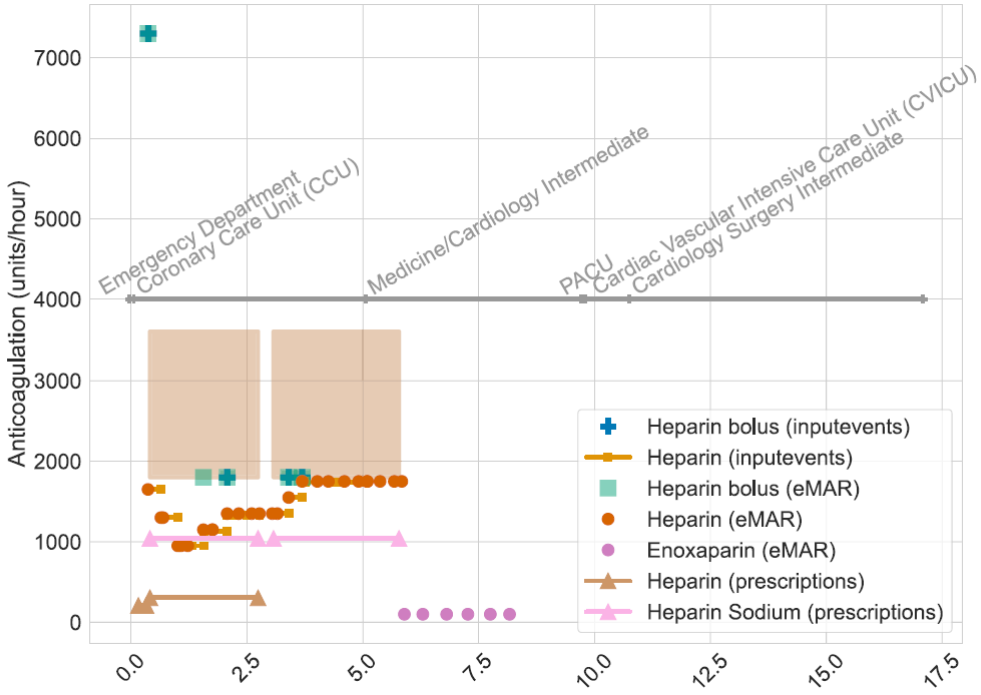
MIMIC IV


bidmc
metavision
external
FORMAT CONVERSION
DEIDENTIFICATION
USER FEEDBACK & CORRECTIONS
MIMIC-IV
HOSPITAL
PATIENT TRACKING
patients
admissions
transfers
ADMINISTRATION
services
poe, poe_detail
BILLING
diagnoses_icd, d_icd_diagnoses
procedures_icd, d_icd_procedures
drgcodes
hcpcsevents, d_hcpcs
MEASUREMENT
microbiologyevents
labevents, d_labitems
omr
MEDICATION
emar, emar_detail
pharmacy
prescriptions
PATIENT TRACKING
icustays
MEASUREMENT
d_items
chartevents
datetimeevents
ingredientevents
inputevents
outputevents
procedureevents
NOTE
DEIDENTIFIED FREE-TEXT
discharge, discharge_detail
radiology, radiology_detail



Visualization of data within MIMIC-IV for a single patient's hospitalization: `hadm_id 28503629`. Three vertically stacked panels highlight the variety of information available. Vital signs are shown in the top panel: note the frequency of data collection for temperature is much higher at the start of the ICU stay due to the use of targeted temperature management. Procedures from multiple sources are shown in the middle panel, including from billing information, the provider order entry system, as well as the ICU information system. The bottom panel displays patient laboratory measurements. Note that while frequent vital signs are only available when the patient is in the ICU, laboratory measures are available throughout their hospitalization.

MIMIC IV Examples



Visualization of medication information documented within MIMIC-IV for a single patient's hospitalization: `hadm_id 28503629`. The annotated grey line indicates care units for the patient throughout their stay. Bolus medications are indicated by markers, continuous infusions as lines, and range doses as filled boxes. For example, on day 5 of their hospital stay, the patient had two active prescriptions for heparin (one for 1600–3500 units of heparin, brown filled box, and one for 1000 units of heparin, pink line with triangles). Additionally on day 5, the patient continued to received heparin according to *emar* (orange circle), and was imminently transferred to the medicine/cardiology intermediate ward.

MIMIC-IV table	Description of change
emar, *emar_detail*, *ingredientevents*, *omr*, *poe*, *poe_detail*, *pharmacy*	New tables only present in MIMIC-IV.
labevents	Added the following columns: `storetime`, `specimen_id`, `ref_range_lower`, `ref_range_upper`, `priority`, and `comments`.
microbiologyevents	Added the following columns: `micro_specimen_id`, `test_seq`, `storedate`, `storetime`, `test_name`, `test_name`, `quantity`, `comments`.
labevents	Added the following columns: `storetime`, `specimen_id`, `ref_range_lower`, `ref_range_upper`, priority, and `comments`.
hcpcsevents, *d_hcpcs*	Replaced the *cptevents* and *d_cpt* tables.
prescriptions	Columns `starttime` and `endtime` replaced `startdate` and `enddate` as the associated time is now available.
	Columns `drug_name_generic` and `drug_name_poe` were removed.
	Columns `pharmacy_id`, form_rx, and `pharmacy_id` were added.
inputevents	Renamed; equivalent to *inputevents_mv* in MIMIC-III. *inputevents_cv* has been removed.
procedureevents	Renamed; equivalent to *procedureevents_mv* in MIMIC-III.
icustays	The unit-level identifier *icustay_id* has been replaced with the general location-based identifier *stay_id*. *stay_id* is used to identify a period of stay within a single location.

Table 2. Major changes between MIMIC-III v1.4 and MIMIC-IV v2.2.

electronic Medicine Administration Record (eMAR)
Online Medical Record (OMR)
the provider order entry (POE) system

Other data at MIMIC

- https://physionet.org/content/

- For example
  - Chest X-ray data: https://physionet.org/content/mimic-cxr/2.0.0/:
  - The MIMIC Chest X-ray (MIMIC-CXR) Database v2.0.0 is a large publicly available dataset of chest radiographs in DICOM format with free-text radiology reports. The dataset contains 377,110 images corresponding to 227,835 radiographic studies performed at the Beth Israel Deaconess Medical Center in Boston, MA. The dataset is de-identified to satisfy the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) Safe Harbor requirements. Protected health information (PHI) has been removed. The dataset is intended to support a wide body of research in medicine including image understanding, natural language processing, and decision support

Other Datasets

- The eICU Collaborative Research Database (eICU-CRD) v2.0 comprises of 200,859 stays at ICUs and step-down units across 208 hospitals in the continental United States - https://www.nature.com/articles/sdata2018178
- The HiRID database contains high-resolution data for almost 34,000 admissions between 2008–2016 at Bern University Hospital in Switzerland - https://physionet.org/content/hirid/1.1.1/
- The Pediatric Intensive Care (PIC) database is sourced from The Children's Hospital at Zhejiang University School of Medicine with 12,881 patients and 13,941 ICU stays admitted from 2010–2018 - http://pic.nbscn.org/

MIMIC III data access

- https://mimic.physionet.org/gettingstarted/access/
  - Start now, it will take 1-2 weeks
  - When you submit the project description, say you are attending this course and will work on MIMIC on course assignments and group projects

- MIMIC Demo Data
  - https://physionet.org/content/mimiciii-demo/1.4/

References

- https://www.nature.com/articles/sdata201635
- https://mimic.physionet.org/
- https://mimic.mit.edu/docs/iii/tables/
- https://physionet.org/content/mimiciv/2.2/
- https://www.nature.com/articles/s41597-022-01899-x