MIMIC-III is a , free database with information about the health of more than 40,000 people who stayed in the critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. The information has been stripped of all personal information. The database has information like demographics, measurements of vital signs taken at the bedside (about one data point per hour), lab test results, procedures, medicines, caregiver notes, imaging reports, and deaths (including those that happened after the patient was released from the hospital).

MIMIC helps with a wide range of analytical studies, including statistics, improving clinical decision rules, and making electronic tools. It is important for three reasons: researchers from anywhere in the world can use it for free, it includes a very large and diverse group of ICU patients, and it has very detailed information like vital signs, lab results, and medications.

Under a data use agreement, MIMIC-III combines de-identified, complete clinical data from patients who were admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. This data is then made available to researchers all over the world. Because the data are public, clinical studies can be repeated and made better in ways that would not be possible otherwise.

The MIMIC-III database was filled with information that was collected during normal hospital care, so it didn't add any extra work for nurses or get in the way of their work. Data was taken from a number of places, including:

Critical care information tools and hospital electronic health record databases have records from the past.

Death by the Social Security Administration The main file.

During the time when the data was collected, there were two different critical care information systems: the Philips CareVue Clinical Information System (types M2331A and M1215A; Philips Health-care, Andover, MA) and the iMDsoft MetaVision ICU (iMDsoft, Needham, MA).

These tools were where clinical information like:

Physiological measures taken by a nurse and time-stamped (for example, recording the heart rate, arterial blood pressure, or breathing rate every hour); care providers' written notes on the patient's progress; and medications and fluid balances given through an intravenous drip.

With the exception of data about fluid intake, which was very different between the CareVue and MetaVision systems in how it was set up, data was combined when the database tables were made. For data that couldn't be combined, a suffix is added to show where the data came from. For example, inputs for patients being monitored with the CareVue system are stored in INPUTEVENTS\_CV, while inputs for patients being monitored with the Metavision system are stored in INPUTEVENTS\_MV. From hospital and lab health record systems, more information was gathered, including:

patient characteristics and deaths in the hospital.

Some examples of lab test results are hematology, chemistry, and microbiology.

discharge summaries and reports of imaging tests and electrocardiograms.

Information about bills, such as International Classification of Diseases, 9th Edition (ICD-9) codes, Diagnosis Related Group (DRG) codes, and Current Procedural Terminology (CPT) codes.

The Social Security Administration's Death Master File was used to find out when people died outside of hospitals.

Before the data was put into the MIMIC-III database, it was deidentified using structured data cleaning and date shifting, as required by the Health Insurance Portability and Accountability Act (HIPAA). For structured data to be deidentified, all 18 of the identifying data elements listed in HIPAA had to be taken out. This included areas like the patient's name, phone number, address, and dates. In particular, times were moved into the future by a random amount for each patient in the same way so that intervals could be kept. This made stays happen between the years 2100 and 2200. Date moving did not change the time of day, the day of the week, or the season. To hide their real ages and follow HIPAA rules, the dates of birth of patients older than 89 were changed. In the database, these patients have ages of over 300 years.

Protected health information was taken out of free-text fields like diagnostic reports and doctor's notes using a thoroughly tested deidentification system based on extensive dictionary look-ups and pattern-matching with regular expressions. As more information is collected, this deidentification device keeps getting more parts.

Institutional Review Boards at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, and the Massachusetts Institute of Technology in Cambridge, Massachusetts, gave their approval to the project. Patients didn't have to give their consent because the project didn't affect professional care and all protected health information was made anonymous.

MIMIC-III is a set of 26 tables that make up a relational database. Identifiers, which usually end in 'ID', are used to link tables together. For example, SUBJECT\_ID is used to identify a unique patient, HADM\_ID is used to identify a unique hospital admission, and ICUSTAY\_ID is used to identify a unique hospital entry to an intensive care unit.

In a set of "events" tables, things like notes, lab tests, and fluid balance are kept track of. For example, the OUTPUTEVENTS table has all the information about a patient's output, while the LABEVENTS table has the results of a patient's lab tests.

Dictionary tables are those that start with "D\_" and give meanings for identifiers. For example, each row of CHARTEVENTS has a single ITEMID that stands for the idea being measured, but it doesn't have the name of the measurement. By joining CHARTEVENTS and D\_ITEMS on ITEMID, you can find out what idea an ITEMID stands for.

When making the MIMIC data model, it was important to find a balance between how easy it was to understand and how close it was to the real world. So, the model is a reflection of the data sources it is based on. The MIMIC database has been changed over time based on user comments. When changes were done, care was taken not to make assumptions about the underlying data. This means that MIMIC-III is a good representation of the raw hospital data.

Five tables, called ADMISSIONS, PATIENTS, ICUSTAYS, SERVICES, and TRANSFERS, are used to describe and keep track of patient stays. Five more tables, D\_CPT, D\_ICD\_DIAGNOSES, D\_ICD\_PROCEDURES, D\_ITEMS, and D\_LABITEMS, are dictionaries that let you look up codes by their meanings. The rest of the tables have information about patient care, like measurements of the patient's body, notes from caregivers, and payment information.

MIMIC-III is given as a set of comma-separated value (CSV) files and tools to help import the data into database systems like PostreSQL, MySQL, and MonetDB. Since the database has a lot of knowledge about how to care for patients clinically, it needs to be treated with care and respect.