UNIVERSIDADE DO VAELO DO RIO DOS SINOS – UNISINOS

A DESCRIPTION OF THE MIMIC-III

São Leopoldo

2016

Summary

A DESCRIPTION OF THE MIMIC-III

Introduction: MIMIC-III (Medical Information Mart for Intensive Care 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 (both in and out of hospital). 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.
- it contains high temporal resolution data including lab results, electronic documentation, and bedside monitor trends and waveforms.

Tables in MIMIC

ADMISSIONS

Overview

Table source: Hospital database.

Table purpose: Define a patient's hospital admission, HADM_ID.

Number of rows: 58976

Links to:

- PATIENTS on SUBJECT_ID

Brief summary

The ADMISSIONS table gives information regarding a patient's admission to the hospital. Since each unique hospital visit for a patient is assigned a unique HADM_ID, the ADMISSIONS table can be considered as a definition table for HADM_ID. Information available includes timing information for admission and discharge, demographic information, the source of the admission, and so on.

Important considerations

- The data is sourced from the admission, discharge and transfer database from the hospital (often referred to as 'ADT' data).
- Organ donor accounts are sometimes created for patients who died in the hospital. These are distinct hospital admissions with very short, sometimes negative lengths of stay. Furthermore, their DEATHTIME is frequently the same as the earlier patient admission's DEATHTIME.
- All text data, except for that in the INSURANCE column, is stored in upper case.

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)

Name	**Postgres data type**
ADMISSION_LOCATION	VARCHAR(50)
DISCHARGE_LOCATION	VARCHAR(50)
INSURANCE	VARCHAR(255)
LANGUAGE	VARCHAR(10)
RELIGION	VARCHAR(50)
MARITAL_STATUS	VARCHAR(50)
ETHNICITY	VARCHAR(200)
EDREGTIME	TIMESTAMP(0)
EDOUTTIME	TIMESTAMP(0)
DIAGNOSIS	VARCHAR(300)
HOSPITAL_EXPIRE_FLAG	TINYINT
HAS_IOEVENTS_DATA	TINYINT
HAS_CHARTEVENTS_DATA	TINYINT

Detailed description

The ADMISSIONS table defines all HADM_ID present in the database, covering an admission period between 1 June 2001 and 10 October 2012.

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.

ADMITTIME, DISCHTIME, DEATHTIME

ADMITTIME provides the date and time the patient was admitted to the hospital, while DISCHTIME provides the date and time the patient was discharged from the hospital. If applicable, DEATHTIME provides the time of in-hospital death for the patient. Note that DEATHTIME is only present if the patient died in-hospital, and is almost always the same as the patient's DISCHTIME. However, there can be some discrepancies due to typographical errors.

ADMISSION_TYPE

ADMISSION_TYPE describes the type of the admission: 'ELECTIVE', 'URGENT', 'NEWBORN' or 'EMERGENCY'. Emergency/urgent indicate unplanned medical care, and are often collapsed into a single category in studies. Elective indicates a previously planned hospital admission. Newborn indicates that the HADM_ID pertains to the patient's birth.

ADMISSION_LOCATION

ADMISSION_LOCATION provides information about the previous location of the patient prior to arriving at the hospital. There are 9 possible values:

- EMERGENCY ROOM ADMIT
- TRANSFER FROM HOSP/EXTRAM
- TRANSFER FROM OTHER HEALT
- CLINIC REFERRAL/PREMATURE
- ** INFO NOT AVAILABLE **
- TRANSFER FROM SKILLED NUR
- TRSF WITHIN THIS FACILITY
- HMO REFERRAL/SICK
- PHYS REFERRAL/NORMAL DELI

The truncated text occurs in the raw data.

INSURANCE, LANGUAGE, RELIGION, MARITAL_STATUS, ETHNICITY

The INSURANCE, LANGUAGE, RELIGION, MARITAL_STATUS, ETHNICITY columns describe patient demographics. These columns occur in the

ADMISSIONS table as they are originally sourced from the admission, discharge, and transfers (ADT) data from the hospital database. The values occasionally change between hospital admissions (HADM_ID) for a single patient (SUBJECT_ID). This is reasonable for some fields (e.g. MARITAL_STATUS, RELIGION), but less reasonable for others (e.g. ETHNICITY).

EDREGTIME, EDOUTTIME

Time that the patient was registered and discharged from the emergency department.

DIAGNOSIS

The DIAGNOSIS column provides a preliminary, free text diagnosis for the patient on hospital admission. The diagnosis is usually assigned by the admitting clinician and does not use a systematic ontology. As of MIMIC-III v1.0 there were 15,693 distinct diagnoses for 58,976 admissions. The diagnoses can be very informative (e.g. chronic kidney failure) or quite vague (e.g. weakness). Final diagnoses for hospital admissions are coded and can be found in the DIAGNOSES_ICD table.

CALLOUT

Overview

Table source: Hospital database.

Table purpose: Provides information when a patient was READY for discharge from the ICU, and when the patient was actually discharged from the ICU.

Number of rows: 34499

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID

Brief summary

The CALLOUT table provides information about ICU discharge planning. When a patient is deemed ready to leave the ICU, they are "called out". This process involves: (i) a care provider registering that the patient is ready to leave the ICU and detailing any specialized precautions required, (ii) a coordinator acknowledging the patient requires a bed outside the ward, (iii) a variable period of time in order to coordinate the transfer, and finally (iv) an outcome: either the patient is called out (discharged) or the call out event is canceled. This table provides information for all of the above.

Important considerations

- Call out data is *not* available for all adult patients, as the data collection only began part way through the collection of the MIMIC database
- Call out data is *never* available for neonates

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
SUBMIT_WARDID	INT
SUBMIT_CAREUNIT	VARCHAR(15)
CURR_WARDID	INT
CURR_CAREUNIT	VARCHAR(15)
CALLOUT_WARDID	INT

Name	Postgres data type
CALLOUT_SERVICE	VARCHAR(10)
REQUEST_TELE	SMALLINT
REQUEST_RESP	SMALLINT
REQUEST_CDIFF	SMALLINT
REQUEST_MRSA	SMALLINT
REQUEST_VRE	SMALLINT
CALLOUT_STATUS	VARCHAR(20)
CALLOUT_OUTCOME	VARCHAR(20)
DISCHARGE_WARDID	INT
ACKNOWLEDGE_STATUS	VARCHAR(20)
CREATETIME	TIMESTAMP(0)
UPDATETIME	TIMESTAMP(0)
ACKNOWLEDGETIME	TIMESTAMP(0)
OUTCOMETIME	TIMESTAMP(0)
FIRSTRESERVATIONTIME	TIMESTAMP(0)
CURRENTRESERVATIONTIME	TIMESTAMP(0)

Detailed description

The CALLOUT table provides information for “call out events”. A call out event occurs as follows: first, a patient is ready for discharge from the ICU. A care provider creates a call out request (occurring at CREATETIME) stating the desired ward and any necessary precautions. The call out request is acknowledged, usually within a short period of time (ACKNOWLEDGETIME).

When the patient is actually transferred out of the ICU, the CALLOUT_OUTCOME becomes 'Discharged' and the OUTCOMETIME column contains the time at which the patient was discharged.

SUBJECT_ID, HADM_ID

SUBJECT_ID and HADM_ID define the patient and hospital admission corresponding to the given call out event. ? Unique

SUBMIT_WARDID, SUBMIT_CAREUNIT

SUBMIT_WARDID identifies the ward from which the request was submitted. SUBMIT_CAREUNIT indicates whether the SUBMIT_WARDID corresponds to an ICU cost center, and if so, what type of ICU cost center.

CURR_WARDID, CURR_CAREUNIT

CURR_WARDID identifies the ward in which the patient resides when called out (i.e. prior to discharge/transfer). CURR_CAREUNIT indicates which ICU cost center the CURR_WARDID corresponds to (note: since all patients are being discharged from an ICU, all patients should reside in an ICU cost center).

CALLOUT_WARDID, CALLOUT_SERVICE

CALLOUT_WARDID identifies the ward to which the patient should be discharged. Note that CALLOUT_WARDID = 0 represents 'Home' and CALLOUT_WARDID = 1 represents 'First available ward'. The remaining IDs correspond to distinct wards in the hospital. CALLOUT_SERVICE is the service under which the patient should be discharged.

REQUEST_TELE, REQUEST_RESP,REQUEST_CDIFF, REQUEST_MRSA,REQUEST_VRE

The request columns are binary indicators which request certain precautions for the patient in the subsequent ward where they will reside. For example, MRSA means that the patient is colonized or infected with MRSA, a drug resistant source of hospital acquired infections.

CALLOUT_STATUS, CALLOUT_OUTCOME

CALLOUT_STATUS indicates whether the call out is still active or not: if a call out is answered it should be flagged as inactive. CALLOUT_OUTCOME is either 'Discharged' or 'Cancelled', indicating whether the patient finally called out (i.e. discharged) or not.

DISCHARGE_WARDID

DISCHARGE_WARDID indicates the ward to which the patient was actually discharged. DISCHARGE_WARDID = 0 indicates home and other values correspond to distinct wards in the hospital.

ACKNOWLEDGE_STATUS

ACKNOWLEDGE_STATUS indicates the response to the callout event: 'Acknowledged', 'Revised', 'Unacknowledged' or 'Reactivated'.

CREATETIME, UPDATETIME,ACKNOWLEDGETIME, OUTCOMETIME,FIRSTRESERVATIONTIME,CURRENTRESERVATIONTIME

CREATETIME provides the time and date that the call out was initiated. UPDATETIME provides the last time at which the call out event was updated. ACKNOWLEDGETIME is the time at which the call out was first acknowledged. OUTCOMETIME is the time at which the CALLOUT_OUTCOME occurred. FIRSTRESERVATIONTIME and CURRENTRESERVATIONTIME provide information regarding ward reservations.

CAREGIVERS

Overview

Table source: CareVue and Metavision ICU databases.

Table purpose: Defines the role of caregivers.

Number of rows: 7567

Links to:

- CHARTEVENTS on CGID

Brief summary

This table provides information regarding care givers. For example, it would define if a care giver is a research nurse (RN), medical doctor (MD), and so on.

Table columns

Name	Postgres data type
ROW_ID	INT
CGID	INT
LABEL	VARCHAR(15)
DESCRIPTION	VARCHAR(30)

Detailed Description

The CAREGIVERS table provides information regarding the type of caregiver. Each caregiver is represented by a unique integer which maps to this table.

CGID

CGID is a unique identifier for each distinct caregiver present in the database. CGID is sourced from two tables in the raw data: the CareVue and Metavision ICU databases. Due to imprecision in the storage of unique identifiers across the database, it is possible that two distinct caregivers with the same names (e.g. RN Sarah Jones and MD Sarah Jones) would be considered as the same caregiver. However, this is an unlikely occurrence.

LABEL

LABEL defines the type of caregiver: e.g. RN, MD, PharmD, etc. Note that LABEL is a free text field and as such contains many typographical errors and spelling variants of the same concept (e.g. MD, MDs, M.D.).

DESCRIPTION

DESCRIPTION is less frequently present than LABEL, and provides additional information regarding the caregiver. This column is much more structured, and contains only 17 unique values as of MIMIC-III v1.0.

CHARTEVENTS

Overview

Table source: CareVue and Metavision ICU databases.

Table purpose: Contains all charted data for all patients.

Number of rows: 263,201,375

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID
- CAREGIVERS on CGID

Brief summary

CHARTEVENTS contains all the charted data available for a patient. During their ICU stay, the primary repository of a patient’s information is their electronic chart. The electronic chart displays patients’ routine vital signs and any additional information relevant to their care: ventilator settings, laboratory values, code status, mental status, and so on. As a result, the bulk of information about a patient's stay is contained in CHARTEVENTS. Furthermore, even though laboratory values are captured elsewhere (LABEVENTS), they are frequently repeated within CHARTEVENTS. This occurs because it is desirable to display the laboratory values on the patient's electronic chart, and so the values are copied from the database storing laboratory values to the database storing the CHARTEVENTS.

Important considerations

- Some items are duplicated between the labevents and chartevents tables. In cases where there is disagreement between measurements, labevents should be taken as the ground truth.

Table columns

Name	Postgres data type	In CareVue	In Metavision
ROW_ID	INT	Y	Y
SUBJECT_ID	NUMBER(7,0)	Y	Y
HADM_ID	NUMBER(7,0)	Y	Y
ICUSTAY_ID	NUMBER(7,0)	Y	Y
ITEMID	NUMBER(7,0)	Y	Y
CHARTTIME	DATE	Y	Y
STORETIME	DATE	Y	Y
CGID	NUMBER(7,0)	Y	Y
VALUE	VARCHAR2(200 BYTE)	Y	Y
VALUENUM	NUMBER	Y	Y

Name	Postgres data type	In CareVue	In Metavision
VALUEUOM	VARCHAR2(20 BYTE)	Y	Y
WARNING	NUMBER(1,0)		Y
ERROR	NUMBER(1,0)		Y
RESULTSTATUS	VARCHAR2(20 BYTE)	Y	
STOPPED	VARCHAR2(20 BYTE)	Y	

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

ITEMID

Identifier for a single measurement type in the database. Each row associated with one ITEMID (e.g. 212) corresponds to an instantiation of the same measurement (e.g. heart rate).

CHARTTIME, STORETIME

CHARTTIME records the time at which an observation was made, and is usually the closest proxy to the time the data was actually measured. STORETIME records the time at which an observation was manually input or manually validated by a member of the clinical staff.

CGID

CGID is the identifier for the caregiver who validated the given measurement.

VALUE, VALUENUM

VALUE contains the value measured for the concept identified by the ITEMID. If this value is numeric, then VALUENUM contains the same data in a numeric format. If this data is not numeric, VALUENUM is null. In some cases (e.g. scores like Glasgow Coma Scale, Richmond Sedation Agitation Scale and Code Status), VALUENUM contains the score and VALUE contains the score and text describing the meaning of the score.

VALUEUOM

VALUEUOM is the unit of measurement for the VALUE, if appropriate.

WARNING, ERROR

WARNING and ERROR are Metavision specific columns which specify if a warning for the value was raised and if an error occurred during the measurement.

RESULTSTATUS, STOPPED

RESULTSTATUS and STOPPED are CareVue specific columns which specify the type of measurement (RESULTSTATUS is ‘Manual’ or ‘Automatic’) and whether the measurement was stopped.

CPTEVENTS

Overview

Table source: Hospital database.

Table purpose: Contains current procedural terminology (CPT) codes, which facilitate billing for procedures performed on patients.

Number of rows: 573146

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID

- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID
- CAREGIVERS on CGID

Brief summary

The CPTEVENTS table contains a list of which current procedural terminology codes were billed for which patients. This can be useful for determining if certain procedures have been performed (e.g. ventilation).

Important considerations

- The respiratory cost center bills for ventilation *regardless* of duration. That means that 30 minutes of mechanical ventilation for a single day would result in same billing code as a full 24 hours of mechanical ventilation.
- Non-invasive ventilation and mechanical ventilation use the same CPT code. Differentiating these concepts requires selecting them using the DESCRIPTION column.

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
COSTCENTER	VARCHAR(10)
CHARTDATE	TIMESTAMP(0)
CPT_CD	VARCHAR(10)
CPT_NUMBER	INT

Name	Postgres data type
CPT_SUFFIX	VARCHAR(5)
TICKET_ID_SEQ	INT
SECTIONHEADER	VARCHAR(50)
SUBSECTIONHEADER	VARCHAR(300)
DESCRIPTION	VARCHAR(200)

Detailed Description

CPTEVENTS contains current procedural terminology (CPT) codes for patients as billed through either the ICU cost center or the respiratory cost center. Each code represents a distinct procedure performed on the patient during their ICU stay.

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

COSTCENTER

COSTCENTER is the cost center which billed for the corresponding CPT codes. There are two possible cost centers: ‘ICU’ and ‘Resp’. ‘Resp’ codes correspond to mechanical or non-invasive ventilation and were billed by the respiratory therapist. ‘ICU’ codes correspond to the procedures billed for by the ICU.

CHARTDATE

The date at which the procedure occurred.

CPT_CD, CPT_NUMBER, CPT_SUFFIX

CPT_CD contains the original CPT code. CPT_NUMBER is a numeric version of the CPT_CD column, which allows for easier range comparisons in querying.

However, note not all CPT_CD are fully numeric. The CPT_SUFFIX column contains the text suffix when the CPT_CD contains non-numeric characters.

TICKET_ID_SEQ

The order of the CPT_CD.

SECTIONHEADER, SUBSECTIONHEADER

The section headers provide a category for the given CPT code. These headers were assigned using the D_CPT table.

DESCRIPTION

In the case of a CPT_CD corresponding to the respiratory cost center, the description provides information about the meaning of the CPT code. Otherwise, the field is null.

DATETIMEEVENTS

Overview

Table source: CareVue and Metavision ICU databases.

Table purpose: Contains all date formatted data.

Number of rows: 4,486,049

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID
- CAREGIVERS on CGID

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ICUSTAY_ID	INT
ITEMID	INT
CHARTTIME	TIMESTAMP(0)
STORETIME	TIMESTAMP(0)
CGID	INT
VALUE	TIMESTAMP(0)
VALUEUOM	VARCHAR(50)
WARNING	SMALLINT
ERROR	SMALLINT
RESULTSTATUS	VARCHAR(50)
STOPPED	VARCHAR(50)

Detailed Description

DATETIMEEVENTS contains all date measurements about a patient in the ICU. For example, the date of last dialysis would be in the DATETIMEEVENTS table, but the systolic blood pressure would not be in this table. As all dates in MIMIC-III are anonymized to protect patient confidentiality, all dates in this table have been shifted. Note that the chronology for an individual patient has been unaffected however, and quantities such as the difference between two dates remain true to reality.

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

ITEMID

CHARTTIME, STORETIME

CHARTTIME records the time at which an observation was charted, and is usually the closest proxy to the time the data was actually measured. STORETIME records the time at which an observation was manually input or manually validated by a member of the clinical staff.

CGID

CGID is the identifier for the caregiver who validated the given measurement.

VALUE

VALUEUOM

WARNING, ERROR

RESULTSTATUS, STOPPED

DIAGNOSES_ICD

Overview

Table source: Hospital database.

Table purpose: Contains ICD diagnoses for patients, most notably ICD-9 diagnoses.

Number of rows: 651,047

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- D_ICD_DIAGNOSES on ICD9_CODE

Important considerations

- All ICD codes in MIMIC-III are ICD-9 based. The Beth Israel Deaconess Medical Center will begin using ICD-10 codes in 2015.
- The code field for the ICD-9-CM Principal and Other Diagnosis Codes is six characters in length, with the decimal point implied between the third and fourth digit for all diagnosis codes other than the V codes. The decimal is implied for V codes between the second and third digit.

Table columns

Name	PostgreSQL data type	Modifiers
ROW_ID	INT	not null
SUBJECT_ID	INT	not null
HADM_ID	INT	not null
SEQ_NUM	INT	
ICD9_CODE	VARCHAR(10)	

Detailed Description

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

SEQ_NUM

SEQ_NUM provides the order in which the ICD diagnoses relate to the patient. ICD diagnoses are ordered by priority - and the order does have an impact on the reimbursement for treatment.

ICD9_CODE

ICD9_CODE contains the actual code corresponding to the diagnosis assigned to the patient for the given row. Note that all codes, as of MIMIC-III v1.0, are ICD-9 codes.

DRGCODES

Overview

Table source: Hospital database.

Table purpose: Contains diagnosis related groups (DRG) codes for patients.

Number of rows: 125,557

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID

Important considerations

- HCFA-DRG codes have multiple descriptions as they have changed over time. Sometimes these descriptions are similar, but sometimes they are *completely different diagnoses*. Users will need to select rows using both the code *and* the description.
- Since there are multiple versions of DRG codes, queries will need to incorporate both the type of DRG and the code when filtering for a certain diagnosis.
- All patients have an HCFA-DRG code, but not all patients have an APR-DRG code. Note that APR-DRG is believed to be an alternative, more specific, code which could be used in conjunction with the HCFA codes.

Table columns

Name	**PostgreSQL data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
DRG_TYPE	VARCHAR(20)
DRG_CODE	VARCHAR(20)
DESCRIPTION	VARCHAR(300)
DRG_SEVERITY	SMALLINT
DRG_MORTALITY	SMALLINT

Detailed Description

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

DRG_TYPE

DRG_TYPE provides the type of DRG code in the entry. There are two types of DRG codes in the database which have overlapping ranges but distinct definitions for the codes. The two types of DRG codes in the MIMIC-III database are 'HCFA' (Health Care Financing Administration) and 'APR' (All Payers Registry).

DRG_CODE

DRG_CODE contains a code which represents the diagnosis billed for by the hospital.

DESCRIPTION

DESCRIPTION provides a human understandable summary of the meaning of the given DRG code. The description field frequently has acronyms which represent comorbidity levels (comorbid conditions or "CC"). The following table provides a definition for some of these acronyms:

Acronym	**Description**
w CC/MCC	with CC or Major CC
w MCC	with Major CC
w CC	with CC and without Major CC
w NonCC	with NonCC and without CC or Major CC
w/o MCC	with CC or Non CC and without Major CC
w/o CC/MCC	with nonCC and without CC or Major CC

Note that there are three levels of comorbidities: none, with comorbid conditions, and with major comorbid conditions. These acronyms are primarily used in HCFA/MS DRG codes.

DRG_SEVERITY, DRG_MORTALITY

DRG_SEVERITY and DRG_MORTALITY provide additional granularity to DRG codes in the 'APR' DRG type. Severity and mortality allow for higher billing costs when a diagnosis is more severe, and vice versa.

D_CPT

Overview

Table source: Online definitions.

Table purpose: High-level definitions for current procedural terminology (CPT) codes.

Number of rows: 134

Links to:

- CPTEVENTS on CPT_CD between MINCODEINSUBSECTION and MAXCODEINSUBSECTION

Brief summary

This table gives some high level information regarding current procedural terminology (CPT) codes. Unfortunately, detailed information for individual codes is unavailable.

Important considerations

- Unlike all other definition tables, D_CPT does *not* have a one to one mapping with the corresponding CPT_CD in CPTEVENTS, rather each row of D_CPT maps to a range of CPT_CD.

Table columns

Name	Postgres data type
ROW_ID	INT
CATEGORY	SMALLINT
SECTIONRANGE	VARCHAR(100)
SECTIONHEADER	VARCHAR(50)
SUBSECTIONRANGE	VARCHAR(100)
SUBSECTIONHEADER	VARCHAR(300)
CODESUFFIX	VARCHAR(5)
MINCODEINSUBSECTION	INT
MAXCODEINSUBSECTION	INT

Detailed Description

D_CPT provides information about CPT codes, specifically it provides the overall purpose of the procedure and in some cases the body system related to the procedure.

CATEGORY

CATEGORY is a integer which identifies the category of the CPT code.

SECTIONRANGE, SECTIONHEADER

SECTIONRANGE defines the range of codes for the given section, and SECTIONHEADER provides the description of the given section. There are 8 possible sections:

- Evaluation and management
- Surgery
- Radiology
- Anesthesia
- Emerging technology
- Pathology and laboratory
- Performance measurement
- Medicine

Note that the evaluation and management tends to represent administrative or generic costs.

SUBSECTIONRANGE, SUBSECTIONHEADER

Similarly as for the sections, SUBSECTIONRANGE defines the range of codes for the given subsection, and SUBSECTIONHEADER provides the description of the given subsection. The subsection provides extra detail which can be useful, for example, when the section header is 'Anesthesia' the subsection provides information on the general anatomical region for the anesthesia.

MINCODEINSUBSECTION, MAXCODEINSUBSECTION

These columns provide numeric representations of minimum and maximum value in the SUBSECTIONRANGE column. This facilitates joining the CPTEVENTS table to the D_CPT table on CPT_CD.

D_ICD_DIAGNOSES

Overview

Table source: Online sources.

Table purpose: Definition table for ICD diagnoses.

Number of rows: 14,567

Links to:

- DIAGNOSES_ICD ON ICD9_CODE

Brief summary

This table defines International Classification of Diseases Version 9 (ICD-9) codes for **diagnoses**. These codes are assigned at the end of the patient's stay and are used by the hospital to bill for care provided.

Table columns

Name	**Postgres data type**
ROW_ID	INT
ICD9_CODE	VARCHAR(10)
SHORT_TITLE	VARCHAR(50)
LONG_TITLE	VARCHAR(300)

Detailed Description

ICD9_CODE

ICD9_CODE is the International Coding Definitions Version 9 (ICD-9) code. Each code corresponds to a single diagnostic concept.

SHORT_TITLE, LONG_TITLE

The title fields provide a brief definition for the given diagnosis code in ICD9_CODE.

D_ICD_PROCEDURES

Overview

Table source: Online sources.

Table purpose: Definition table for ICD procedures.

Number of rows: 3,882

Links to:

- PROCEDURES_ICD on ICD9_CODE

Brief summary

This table defines International Classification of Diseases Version 9 (ICD-9) codes for **procedures**. These codes are assigned at the end of the patient's stay and are used by the hospital to bill for care provided. They can further be used to identify if certain procedures have been performed (e.g. surgery).

Table columns

Name	Postgres data type
ROW_ID	INT
ICD9_CODE	VARCHAR(10)
SHORT_TITLE	VARCHAR(50)
LONG_TITLE	VARCHAR(300)

Detailed Description

ICD9_CODE

ICD9_CODE is the International Coding Definitions Version 9 (ICD-9) code. Each code corresponds to a single procedural concept.

SHORT_TITLE, LONG_TITLE

The title fields provide a brief definition for the given procedure code in ICD9_CODE.

D_ITEMS

Overview

Table source: CareVue and Metavision ICU databases.

Table purpose: Definition table for all items in the ICU databases.

Number of rows: 12,478

Links to:

- CHARTEVENTS on ITEMID
- DATETIMEEVENTS on ITEMID

- INPUTEVENTS_CV on ITEMID
- INPUTEVENTS_MV on ITEMID
- MICROBIOLOGYEVENTS on ITEMID
- OUTPUTEVENTS on ITEMID
- PROCEDUREEVENTS_MV on ITEMID

Important considerations

- D_ITEMS is sourced from two *distinct* ICU databases. The main consequence is that there are duplicate ITEMID for each concept. For example, heart rate is captured both as an ITEMID of 212 (CareVue) and as an ITEMID of 220045 (Metavision). As a result, it is necessary to search for multiple ITEMID to capture a single concept across the entire database. This can be tedious, and it is an active project to coalesce these ITEMID - one which welcomes any and all help provided by the community!
- Another source of duplicate ITEMID is due to the free text nature of data entry in CareVue - as a result there are additional ITEMID which correspond to misspellings or synonymous descriptions of a single concept. It is important to search for all possible abbreviations and descriptions of a concept to capture all associated ITEMID.
- If the LINKSTO column is null, then the data is currently unavailable, but planned for a future release.

Table columns

Name	**Postgres data type**
ROW_ID	INT
ITEMID	INT
LABEL	VARCHAR(200)

Name	**Postgres data type**
ABBREVIATION	VARCHAR(100)
DBSOURCE	VARCHAR(20)
LINKSTO	VARCHAR(50)
CATEGORY	VARCHAR(100)
UNITNAME	VARCHAR(100)
PARAM_TYPE	VARCHAR(30)
CONCEPTID	INT

Detailed Description

The D_ITEMS table defines ITEMID, which represents measurements in the database. Measurements of the same type (e.g. heart rate) will have the same ITEMID (e.g. 211). The ITEMID column is an alternate primary key to this table: it is unique to each row.

Note that the D_ITEMS table is sourced from two ICU databases: Metavision and CareVue. Each system had its own set of ITEMID to identify concepts. As a result, there are multiple ITEMID which correspond to the same concept. For CareVue data, ITEMID = 211 is used to identify heart rates, whereas for Metavision data, ITEMID = 220045 is used. All Metavision ITEMIDs will have a value > 220000.

Note that the D_ITEMS table does *not* link to the LABEVENTS table, as this data was acquired separately from the hospital database. The D_ITEMS table was acquired from the ICU databases.

ITEMID

As an alternate primary key to the table, ITEMID is unique to each row.

LABEL, ABBREVIATION

The LABEL column describes the concept which is represented by the ITEMID. The ABBREVIATION column, only available in Metavision, lists a common abbreviation for the label.

DBSOURCE

The DBSOURCE column was generated to clarify which database the given ITEMID was sourced from: ‘carevue’ indicates the ITEMID was sourced from CareVue, while ‘metavision’ indicated the ITEMID was sourced from Metavision.

LINKSTO

LINKSTO provides the table name which the data links to. For example, a value of ‘chartevents’ indicates that the ITEMID of the given row is contained in CHARTEVENTS. A single ITEMID is only used in one event table, that is, if an ITEMID is contained in CHARTEVENTS it will *not* be contained in any other event table (e.g. IOEVENTS, CHARTEVENTS, etc).

CATEGORY

CATEGORY provides some information of the type of data the ITEMID corresponds to. Examples include ‘ABG’, which indicates the measurement is sourced from an arterial blood gas, ‘IV Medication’, which indicates that the medication is administered through an intravenous line, and so on.

UNITNAME

UNITNAME specifies the unit of measurement used for the ITEMID. This column is not always available, and this may be because the unit of measurement varies, a unit of measurement does not make sense for the given data type, or the unit of measurement is simply missing. Note that there is sometimes additional information on the unit of measurement in the associated event table, e.g. the VALUEUOM column in CHARTEVENTS.

PARAM_TYPE

PARAM_TYPE describes the type of data which is recorded: a date, a number or a text field.

D_LABITEMS

Overview

Table source: Hospital database.

Table purpose: Definition table for all laboratory measurements.

Number of rows: 755

Links to:

- LABEVENTS on ITEMID

Important considerations

- The ITEMID from MIMIC-III v1.0 does *not* match the ITEMID from MIMIC-II v2.6. If a mapping between the two is necessary, please contact the guardians of the database.
- Many of the LOINC codes were assigned during a project to standardize the ontology of lab measurements in the MIMIC database. Consequently, the codes were assigned post-hoc, and may not be present for every lab measurement. We welcome improvements to the present codes or assignment of LOINC codes to unmapped data elements from the community.

Table columns

Name	**Postgres data type**
ROW_ID	INT
ITEMID	INT
LABEL	VARCHAR(100)

Name	Postgres data type
FLUID	VARCHAR(100)
CATEGORY	VARCHAR(100)
LOINC_CODE	VARCHAR(100)

Detailed Description

D_LABITEMS contains definitions for all ITEMID associated with lab measurements in the MIMIC database. All data in LABEVENTS link to the D_LABITEMS table. Each unique LABEL in the hospital database was assigned an ITEMID in this table, and the use of this ITEMID facilitates efficient storage and querying of the data. Note that lab items are kept separate while most definitions are contained in the D_ITEMS table, and there were good reasons to keep the lab items separate.

As the laboratory data is acquired from the hospital database, the data is consistent across all years in the database. Consequently, there is usually only one ITEMID associated with each concept in the database. Furthermore, the data contains information collected in departments outside the ICU. This includes both wards within the hospital and clinics outside the hospital. Most concepts in this table have been mapped to LOINC codes, an openly available ontology which provides a rich amount of information about the laboratory measurement including reference ranges, common units of measurement and other further detail regarding the measurement.

ITEMID

As a candidate key in the table, ITEMID is unique to each row.

LABEL

The LABEL column describes the concept which is represented by the ITEMID.

FLUID

FLUID describes the substance on which the measurement was made. For example, chemistry measurements are frequently performed on blood, which is listed in this column as ‘BLOOD’. Many of these measurements are also acquirable on other fluids, such as urine, and this column differentiates these distinct concepts.

CATEGORY

CATEGORY provides higher level information as to the type of measurement. For example, a category of ‘ABG’ indicates that the measurement is an arterial blood gas.

LOINC_CODE

LOINC_CODE contains the LOINC code associated with the given ITEMID. LOINC is an ontology which originally specified laboratory measurements but has since expanded to cover a wide range of clinically relevant concepts. LOINC openly provide a table which contains a large amount of detail about each LOINC code. This table is freely available online or can be provided by the guardians of the database.

ICUSTAYS

Overview

Table source: Hospital database.

Table purpose: Defines each ICUSTAY_ID in the database, i.e. defines a single ICU stay.

Number of rows: 61,532

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID

Important considerations

- ICUSTAY_ID is a *generated* identifier that is *not* based on any raw data identifier. The hospital and ICU databases are not intrinsically linked and so do not have any concept of an ICU encounter identifier.

- The ICUSTAYS table is derived from the TRANSFERS table. Specifically, it groups the TRANSFERS table based on ICUSTAY_ID, and excludes rows where no ICUSTAY_ID is present.

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ICUSTAY_ID	INT
DBSOURCE	VARCHAR(20)
FIRST_CAREUNIT	VARCHAR(20)
LAST_CAREUNIT	VARCHAR(20)
FIRST_WARDID	SMALLINT
LAST_WARDID	SMALLINT
INTIME	TIMESTAMP(0)
OUTTIME	TIMESTAMP(0)
LOS	DOUBLE

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

DBSOURCE

DBSOURCE contains the original ICU database the data was sourced from. Patients admitted between 2001 - 2008 had their data managed by the CareVue information system, represented in this column as ‘carevue’. Patients admitted between 2008 - 2012 had their data managed by the Metavision system, represented in this column as ‘metavision’. Knowing the database source is important as the data archiving for these two databases differs in some cases.

FIRST_CAREUNIT, LAST_CAREUNIT

FIRST_CAREUNIT and LAST_CAREUNIT contain, respectively, the first and last ICU type in which the patient was cared for. As an ICUSTAY_ID groups all ICU admissions within 24 hours of each other, it is possible for a patient to be transferred from one type of ICU to another and have the same ICUSTAY_ID.

Care units are derived from the TRANSFERS table, and definition for the abbreviations can be found in the documentation for TRANSFERS.

FIRST_WARDID, LAST_WARDID

FIRST_WARDID and LAST_WARDID contain the first and last ICU unit in which the patient stayed. Note the grouping of physical locations in the hospital database is referred to as ward. Though in practice ICUs are not referred to as wards, the hospital database technically tracks ICUs as “wards with an ICU cost center”. As a result, each ICU is associated with a WARDID.

INTIME, OUTTIME

INTIME provides the date and time the patient was transferred into the ICU. OUTTIME provides the date and time the patient was transferred out of the ICU.

LOS

LOS is the length of stay for the patient for the given ICU stay, which may include one or more ICU units.

INPUTEVENTS_CV

Overview

Table source: CareVue ICU databases.

Table purpose: Input data for patients.

Number of rows: 17,528,894

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID
- CAREGIVERS on CGID

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ICUSTAY_ID	INT
CHARTTIME	TIMESTAMP(0)
ITEMID	INT
AMOUNT	DOUBLE PRECISION
AMOUNTUOM	VARCHAR(30)
RATE	DOUBLE PRECISION

Name	Postgres data type
RATEUOM	VARCHAR(30)
STORETIME	TIMESTAMP(0)
CGID	BIGINT
ORDERID	BIGINT
LINKORDERID	BIGINT
STOPPED	VARCHAR(30)
NEWBOTTLE	INT
ORIGINALAMOUNT	DOUBLE PRECISION
ORIGINALAMOUNTUOM	VARCHAR(30)
ORIGINALROUTE	VARCHAR(30)
ORIGINALRATE	DOUBLE PRECISION
ORIGINALRATEUOM	VARCHAR(30)
ORIGINALSITE	VARCHAR(30)

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

STARTTIME, ENDTIME

STARTTIME and ENDTIME record the start and end time of an input/output event.

- STARTTIME is only available for inputs, and is unavailable for a large portion of the data (see the understanding IOEVENTS page)
- ENDTIME is the only time used for outputs

ITEMID

Identifier for a single measurement type in the database. Each row associated with one ITEMID (e.g. 212) corresponds to an instantiation of the same measurement (e.g. heart rate). Metavision ITEMID values are all above 220000. A subset of commonly used medications in CareVue data have ITEMID values are between 30000-39999. The remaining input/output ITEMID values are between 40000-49999.

AMOUNT, AMOUNTUOM

AMOUNT and AMOUNTUOM list the amount of a drug or substance administered to the patient either between the STARTTIME and ENDTIME (if both are available) or at the ENDTIME (when the exact start time is unknown, but usually up to an hour before).

RATE, RATEUOM

RATE and RATEUOM list the rate at which the drug or substance was administered to the patient either between the STARTTIME and ENDTIME (if both are available), or it lists the rate at which the drug is *currently* administered at the ENDTIME.

STORETIME

STORETIME records the time at which an observation was manually input or manually validated by a member of the clinical staff.

CGID

CGID is the identifier for the caregiver who validated the given measurement.

ORDERID, LINKORDERID

ORDERID links multiple items contained in the same solution together. For example, when a solution of noradrenaline and normal saline is administered both noradrenaline and normal saline occur on distinct rows but will have the same ORDERID.

LINKORDERID links the same order across multiple instantiations: for example, if the rate of delivery for the solution with noradrenaline and normal saline is changed, two new rows which share the same new ORDERID will be generated, but the LINKORDERID will be the same.

STOPPED, NEWBOTTLE

STOPPED indicates whether the infusion has been disconnected or continued. NEWBOTTLE indicates if a new preparation of the solution was hung at the bedside.

ORIGINALAMOUNT, ORIGINALAMOUNTUOM, ORIGINALROUTE, ORIGINALRATE, ORIGINALRATEUOM, ORIGINALSITE

These columns provide information about the solution the medication was a part of when it was first entered into the information system.

INPUTEVENTS_MV

Overview

Table source: Metavision ICU databases.

Table purpose: Input data for patients.

Number of rows: 3,618,991

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID
- CAREGIVERS on CGID

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ICUSTAY_ID	INT
STARTTIME	TIMESTAMP(0)
ENDTIME	TIMESTAMP(0)
ITEMID	INT
AMOUNT	DOUBLE PRECISION
AMOUNTUOM	VARCHAR(30)
RATE	DOUBLE PRECISION
RATEUOM	VARCHAR(30)
STORETIME	TIMESTAMP(0)
CGID	BIGINT
ORDERID	BIGINT
LINKORDERID	BIGINT
ORDERCATEGORYNAME	VARCHAR(100)
SECONDARYORDERCATEGORYNAME	VARCHAR(100)
ORDERCOMPONENTTYPEDESCRIPTION	VARCHAR(200)
ORDERCATEGORYDESCRIPTION	VARCHAR(50)
PATIENTWEIGHT	DOUBLE PRECISION

Name	**Postgres data type**
TOTALAMOUNT	DOUBLE PRECISION
TOTALAMOUNTUOM	VARCHAR(50)
ISOPENBAG	SMALLINT
CONTINUEINNEXTDEPT	SMALLINT
CANCELREASON	SMALLINT
STATUSDESCRIPTION	VARCHAR(30)
COMMENTS_STATUS	VARCHAR(30)
COMMENTS_TITLE	VARCHAR(100)
COMMENTS_DATE	TIMESTAMP(0)
ORIGINALAMOUNT	DOUBLE PRECISION
ORIGINALRATE	DOUBLE PRECISION

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

STARTTIME, ENDTIME

STARTTIME and ENDTIME record the start and end time of an input/output event.

- STARTTIME is only available for inputs, and is unavailable for a large portion of the data
- ENDTIME is the only time used for outputs

ITEMID

Identifier for a single measurement type in the database. Each row associated with one ITEMID (e.g. 212) corresponds to an instantiation of the same measurement (e.g. heart rate). Metavision ITEMID values are all above 220000. A subset of commonly used medications in CareVue data have ITEMID values are between 30000-39999. The remaining input/output ITEMID values are between 40000-49999.

AMOUNT, AMOUNTUOM

AMOUNT and AMOUNTUOM list the amount of a drug or substance administered to the patient either between the STARTTIME and ENDTIME (if both are available) or at the ENDTIME (when the exact start time is unknown, but usually up to an hour before).

RATE, RATEUOM

RATE and RATEUOM list the rate at which the drug or substance was administered to the patient either between the STARTTIME and ENDTIME (if both are available), or it lists the rate at which the drug is *currently* administered at the ENDTIME.

STORETIME

STORETIME records the time at which an observation was manually input or manually validated by a member of the clinical staff.

CGID

CGID is the identifier for the caregiver who validated the given measurement.

ORDERID, LINKORDERID

ORDERID links multiple items contained in the same solution together. For example, when a solution of noradrenaline and normal saline is administered both noradrenaline and normal saline occur on distinct rows but will have the same ORDERID.

LINKORDERID links the same order across multiple instantiations: for example, if the rate of delivery for the solution with noradrenaline and normal saline is changed, two new rows which share the same new ORDERID will be generated, but the LINKORDERID will be the same.

ORDERCATEGORYNAME, SECONDARYORDERCATEGORYNAME, ORDERCOMPONENTTYPEDESCRIPTION, ORDERCATEGORYDESCRIPTION

These columns provide higher level information about the order the medication/solution is a part of. Categories represent the type of administration, while the ORDERCOMPONENTTYPEDESCRIPTION describes the role of the substance in the solution (i.e. main order parameter, additive, or mixed solution)

PATIENTWEIGHT

The patient weight in kilograms.

TOTALAMOUNT, TOTALAMOUNTUOM

The total amount of the substance in the bag containing the solution.

STATUSDESCRIPTION

STATUSDESCRIPTION states the ultimate status of the item. ‘Stopped’ indicates that the caregiver stopped the item or the programmed volume came to an end. ‘Finished running’ indicates that the programmed volume has come to an end. ‘Rewritten’ indicates that the caregiver rewrote the item, for example making an amendment to the starttime. ‘Changed’ indicates that the caregiver changed an item, for example setting a new rate or dose. Flushed’ indicates that a line was flushed.

ISOPENBAG

Whether the order was from an open bag.

CONTINUEINNEXTDEPT

If the order ended on patient transfer, this field indicates if it continued into the next department (e.g. a floor).

CANCELREASON

If the order was canceled, this column provides some explanation.

COMMENTS_STATUS, COMMENTS_TITLE, COMMENTS_DATE

Specifies if the order was edited or canceled, and if so, the date and job title of the care giver who canceled or edited it.

ORIGINALAMOUNT, ORIGINALRATE

Information on the initial amount and rate of the order.

LABEVENTS

Overview

Table source: Hospital database.

Table purpose: Contains all laboratory measurements for a given patient, including out patient data.

Number of rows: 27,872,575

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- D_LABITEMS on ITEMID

Important considerations

- Note that the time associated with this result is the time of the fluid *acquisition*, not the time that the values were made available to the clinical staff.
- The labevents table contains both in-hospital laboratory measurements *and* out of hospital laboratory measurements from clinics which the patient has visited (since the patient is not "in" a hospital when visiting a clinic, these patients often referred to as "out patients" and the

data is often called “out patient” data). Laboratory measurements for out patients **does not have a HADM_ID**.

- In MIMIC-III v1.0, there is a subset of patients for which the outpatient lab data is not available. They can be identified by checking for patients whose data *always* has an HADM_ID.
- In MIMIC-III v1.0, there is a subset of patients for which text laboratory data is missing. This primarily affects the blood gas type recorded with blood gases.
- Some items are duplicated between the labevents and chartevents tables. In cases where there is disagreement between measurements, labevents should be taken as the ground truth.

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ITEMID	INT
CHARTTIME	TIMESTAMP(0)
VALUE	VARCHAR(200)
VALUENUM	DOUBLE PRECISION
VALUEUOM	VARCHAR(20)
FLAG	VARCHAR(20)

Detailed Description

The LABEVENTS data contains information regarding laboratory based measurements. The process for acquiring a lab measurement is as follows: first, a member of the clinical staff acquires a fluid from a site in the patient's body (e.g. blood from an arterial line, urine from a catheter, etc). Next, the fluid is bar coded to associate it with the patient *and* timestamped to record the time of the fluid acquisition. The lab analyses the data and returns a result within 4-12 hours.

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

ITEMID

Identifier for a single measurement type in the database. Each row associated with one ITEMID (e.g. 212) corresponds to an instantiation of the same measurement (e.g. heart rate).

CHARTTIME

CHARTTIME records the time at which an observation was charted, and is usually the closest proxy to the time the data was actually measured. Note that because the data is directly sourced from the laboratory database, it is *not* validated by ICU clinical staff, and as a result there is no associated STORETIME.

VALUE, VALUENUM

VALUE contains the value measured for the concept identified by the ITEMID. If this value is numeric, then VALUENUM contains the same data in a numeric format. If this data is not numeric, VALUENUM is null. In some cases (e.g. scores like Glasgow Coma Scale, Richmond Sedation Agitation Scale and Code Status), VALUENUM contains the score and VALUE contains the score and text describing the meaning of the score.

VALUEUOM

VALUEUOM is the unit of measurement for the VALUE, if appropriate.

FLAG

FLAG indicates whether the laboratory value is considered abnormal or not, using pre-defined thresholds.

MICROBIOLIGYEVENTS

Overview

Table source: Hospital database.

Table purpose: Contains microbiology information, including tests performed and sensitivities.

Number of rows: 328,446

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- D_ITEMS on SPEC_ITEMID
- D_ITEMS on ORG_ITEMID
- D_ITEMS on AB_ITEMID

Important considerations

- The MICROBIOLOGYEVENTS table does not contain cultures from samples taken outside the ICU
- If the specimen is null, then the culture had no growth reported.

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
CHARTDATE	TIMESTAMP(0)
CHARTTIME	TIMESTAMP(0)
SPEC_ITEMID	INT
SPEC_TYPE_CD	VARCHAR(10)
SPEC_TYPE_DESC	VARCHAR(100)
ORG_ITEMID	INT
ORG_CD	INT
ORG_NAME	VARCHAR(100)
ISOLATE_NUM	SMALLINT
AB_ITEMID	INT
AB_CD	INT
AB_NAME	VARCHAR(30)
DILUTION_TEXT	VARCHAR(10)
DILUTION_COMPARISON	VARCHAR(20)
DILUTION_VALUE	DOUBLE PRECISION
INTERPRETATION	VARCHAR(5)

Detailed Description

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

CHARTDATE, CHARTTIME

CHARTTIME records the time at which an observation was charted, and is usually the closest proxy to the time the data was actually measured. CHARTDATE is the same as CHARTTIME, except there is no time available.

CHARTDATE was included as time information is not always available for microbiology measurements: in order to be clear about when this occurs, CHARTTIME is null, and CHARTDATE contains the date of the measurement.

In the cases where both CHARTTIME and CHARTDATE exists, CHARTDATE is equal to a truncated version of CHARTTIME (i.e. CHARTTIME without the timing information). Not all observations have a CHARTTIME, but all observations have a CHARTDATE.

SPEC_ITEMID, SPEC_TYPE_CD, SPEC_TYPE_DESC

Details the itemid, code, and description for the specimen.

ORG_ITEMID, ORG_CD, ORG_NAME

ISOLATE_NUM

AB_ITEMID, AB_CD, AB_NAME

DILUTION_TEXT, DILUTION_COMPARISON, DILUTION_VALUE

INTERPRETATION

INTERPRETATION indicates the results of the test. “S” is sensitive, “R” is resistant, “I” is intermediate, and “P” is pending.

NOTEEVENTS

Overview

Table source: Hospital database.

Table purpose: Contains all notes for patients.

Number of rows:

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- CAREGIVERS on CGID

Important considerations

- TEXT is often large and contains many newline characters: it may be easier to read if viewed in a distinct program rather than the one performing the queries.
- Echo reports, ECG reports, and radiology reports are available for both inpatient and outpatient stays. If a patient is an outpatient, there will not be an HADM_ID associated with the note. If the patient is an inpatient, but was not admitted to the ICU for that particular hospital admission, then there will *not* be an HADM_ID associated with the note.
- Echos are generated using templates and in some cases there may be discrepancies in severity. For example one report may contain: “Mild PA systolic hypertension. Severe PA systolic hypertension.” indicating that the caregiver may not have removed the appropriate item from the template.

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
CHARTDATE	TIMESTAMP(0)
CATEGORY	VARCHAR(50)
DESCRIPTION	VARCHAR(300)
CGID	INT
ISERROR	CHAR(1)
TEXT	TEXT

Detailed Description

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

CHARTDATE

CHARTDATE records the date at which the note was charted.

CATEGORY, DESCRIPTION

CATEGORY and DESCRIPTION define the type of note recorded. For example, a CATEGORY of ‘Discharge’ indicates that the note is a discharge note, and a DESCRIPTION of ‘Summary’ in conjunction with this indicates that the note is a discharge summary.

CGID

CGID is the identifier for the caregiver who input the note.

ISERROR

TEXT

TEXT contains the note text.

OUTPUTEVENTS

Overview

Table source: CareVue and Metavision ICU databases.

Table purpose: Output data for patients.

Number of rows: 4,349,339

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID
- CAREGIVERS on CGID

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT

Name	Postgres data type
ICUSTAY_ID	INT
CHARTTIME	TIMESTAMP(0)
ITEMID	INT
VALUE	DOUBLE PRECISION
VALUEUOM	VARCHAR(30)
STORETIME	TIMESTAMP(0)
CGID	BIGINT
STOPPED	VARCHAR(30)
NEWBOTTLE	INT
ISERROR	SMALLINT

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

CHARTTIME

CHARTTIME is the time of an output event.

ITEMID

Identifier for a single measurement type in the database. Each row associated with one ITEMID (e.g. 212) corresponds to an instantiation of the same measurement (e.g. heart rate).

Metavision ITEMID values are all above 220000. A subset of commonly used medications in CareVue data have ITEMID values are between 30000-39999. The remaining input/output ITEMID values are between 40000-49999.

VALUE, VALUEUOM

VALUE and VALUEUOM list the amount of a substance at the CHARTTIME (when the exact start time is unknown, but usually up to an hour before).

STORETIME

STORETIME records the time at which an observation was manually input or manually validated by a member of the clinical staff.

CGID

CGID is the identifier for the caregiver who validated the given measurement.

STOPPED, NEWBOTTLE, ISERROR

STOPPED indicates if the order was disconnected at the given CHARTTIME. NEWBOTTLE indicates that a new bag of solution was hung at the given CHARTTIME. ISERROR is a Metavision checkbox where a care giver can specify that an observation is an error. No other details are provided.

PATIENTS

Overview

Table source: CareVue and Metavision ICU databases.

Table purpose: Contains all charted data for all patients.

Number of rows: 46,520

Links to:

- ADMISSIONS on SUBJECT_ID
- ICUSTAYS on SUBJECT_ID

Important considerations

- DOB has been shifted for patients older than 89. The median age for the patients whose date of birth was shifted is 91.4.

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
GENDER	VARCHAR(5)
DOB	TIMESTAMP(0)
DOD	TIMESTAMP(0)
DOD_HOSP	TIMESTAMP(0)
DOD_SSN	TIMESTAMP(0)
EXPIRE_FLAG	VARCHAR(5)

Detailed Description

SUBJECT_ID

SUBJECT_ID is a unique identifier which specifies an individual patient. SUBJECT_ID is a candidate key for the table, so is unique for each row. Information that is consistent for the lifetime of a patient is stored in this table.

GENDER

GENDER is the genotypical sex of the patient.

DOB

DOB is the date of birth of the given patient. Patients who are older than 89 years old at any time in the database have had their date of birth shifted to obscure their age and comply with HIPAA. The shift process was as follows: the patient's age at their first admission was determined. The date of birth was then set to exactly 300 years before their first admission.

DOD, DOD_HOSP, DOD_SSN

DOD is the date of death for the given patient. DOD_HOSP is the date of death as recorded in the hospital database. DOD_SSN is the date of death from the social security database. Note that DOD merged together DOD_HOSP and DOD_SSN, giving priority to DOD_HOSP if both were recorded.

EXPIRE_FLAG

EXPIRE_FLAG is a binary flag which indicates whether the patient died, i.e. whether DOD is null or not. These deaths include both deaths within the hospital (DOD_HOSP) and deaths identified by matching the patient to the social security master death index (DOD_SSN).

PRESCRIPTIONS

Overview

Table source: Hospital provider order entry database.

Table purpose: Contains medication related order entries, i.e. prescriptions.

Number of rows: 4,156,848

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID

Important considerations

- The table does *not* specify if an order was later cancelled as of MIMIC-III v1.0.

Table columns

Name	**Postgres data type**
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ICUSTAY_ID	INT
STARTTIME	TIMESTAMP(0)
ENDTIME	TIMESTAMP(0)
DRUG_TYPE	VARCHAR(100)
DRUG	VARCHAR(100)
DRUG_NAME_POE	VARCHAR(100)
DRUG_NAME_GENERIC	VARCHAR(100)
FORMULARY_DRUG_CD	VARCHAR(120)
GSN	VARCHAR(200)
NDC	VARCHAR(120)
PROD_STRENGTH	VARCHAR(120)
DOSE_VAL_RX	VARCHAR(120)
DOSE_UNIT_RX	VARCHAR(120)
FORM_VAL_DISP	VARCHAR(120)

Name	Postgres data type
FORM_UNIT_DISP	VARCHAR(120)
ROUTE	VARCHAR(120)

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

STARTTIME, ENDTIME

STARTTIME and ENDTIME specify the time period for which the prescription was valid.

DRUG_TYPE

DRUG_TYPE provides the type of drug prescribed.

DRUG, DRUG_NAME_POE,DRUG_NAME_GENERIC

These columns are various representations of the drug prescribed to the patient.

FORMULARY_DRUG_CD, GSN, NDC

These columns provide a representation of the drug in various coding systems. GSN is the Generic Sequence Number. NDC is the National Drug Code.

PROD_STRENGTH

DOSE_VAL_RX, DOSE_UNIT_RX

FORM_VAL_DISP, FORM_UNIT_DISP

ROUTE

The route prescribed for the drug.

PROCEDUREEVENTS_MV

Overview

Table source: Metavision ICU database.

Table purpose: Contains procedures for patients

Number of rows: 258,066

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID
- D_ITEMS on ITEMID

Table columns

Name	Postgres data type
ROW_ID	INT NOT NULL
SUBJECT_ID	INT NOT NULL
HADM_ID	INT NOT NULL
ICUSTAY_ID	INT
STARTTIME	TIMESTAMP(0)
ENDTIME	TIMESTAMP(0)
ITEMID	INT
VALUE	DOUBLE PRECISION

Name	**Postgres data type**
VALUEUOM	VARCHAR(30)
LOCATION	VARCHAR(30)
LOCATIONCATEGORY	VARCHAR(30)
STORETIME	TIMESTAMP(0)
CGID	INT
ORDERID	INT
LINKORDERID	INT
ORDERCATEGORYNAME	VARCHAR(100)
SECONDARYORDERCATEGORYNAME	VARCHAR(100)
ORDERCATEGORYDESCRIPTION	VARCHAR(50)
ISOPENBAG	SMALLINT
CONTINUEINNEXTDEPT	SMALLINT
CANCELREASON	SMALLINT
STATUSDESCRIPTION	VARCHAR(30)
COMMENTS_EDITEDBY	VARCHAR(30)
COMMENTS_CANCELEDBY	VARCHAR(30)
COMMENTS_DATE	TIMESTAMP(0)

PROCEDURES_ICD

Overview

Table source: Hospital database.

Table purpose: Contains ICD procedures for patients, most notably ICD-9 procedures.

Number of rows: 240,095

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- D_ICD_PROCEDURES on ICD9_CODE

Important considerations

- In MIMIC-III v1.0, only ICD-9 codes are used for recording procedures.

Table columns

Name	**PostgreSQL data type**	**Modifiers**
ROW_ID	INT	not null
SUBJECT_ID	INT	not null
HADM_ID	INT	not null
SEQ_NUM	INT	
ICD9_CODE	VARCHAR(10)	

Detailed Description

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

SEQ_NUM

PROC_SEQ_NUM provides the order in which the procedures were performed.

ICD9_CODE

CODE provides the code for the given procedure.

SERVICES

Overview

Table source: Hospital database.

Table purpose: Lists services that a patient was admitted/transferred under.

Number of rows: 73,343

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID

Brief summary

The services table describes the service that a patient was admitted under. While a patient can be physicially located at a given ICU type (say MICU), they are not necessarily being cared for by the team which staffs the MICU. This can happen due to a number of reasons, including bed shortage. The SERVICES table should be used if interested in identifying the type of service a patient is receiving in the hospital. For example, if interested in identifying surgical patients, the recommended method is searching for patients admitted under a surgical service.

Each service is listed in the table as an abbreviation - this is exactly how the data is stored in the hospital database. For user convenience, we have provided a description of each service type.

Service	**Description**
CMED	Cardiac Medical - for non-surgical cardiac related admissions
CSURG	Cardiac Surgery - for surgical cardiac admissions
DENT	Dental - for dental/jaw related admissions
ENT	Ear, nose, and throat - conditions primarily affecting these areas
GU	Genitourinary - reproductive organs/urinary system
GYN	Gynecological - female reproductive systems and breasts
MED	Medical - general service for internal medicine
NB	Newborn - infants born at the hospital
NBB	Newborn baby - infants born at the hospital
NMED	Neurologic Medical - non-surgical, relating to the brain
NSURG	Neurologic Surgical - surgical, relating to the brain
OBS	Observation - not ill enough for a service but kept in hospital for observation
ORTHO	Orthopaedic - surgical, relating to the musculoskeletal system
OMED	Orthopaedic medicine - non-surgical, relating to musculoskeletal system
PSURG	Plastic - restortation/reconstruction of the human body (including cosmetic or aesthetic)
PSYCH	Psychiatric - mental disorders relating to mood, behaviour, cognition, or perceptions
SURG	Surgical - general surgical service not classified elsewhere
TRAUM	Trauma - injury or damage caused by physical harm from an external source

Service	Description
TSURG	Thoracic Surgical - surgery on the thorax, located between the neck and the abdomen
VSURG	Vascular Surgical - surgery relating to the circulatory system

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
TRANSFERTIME	TIMESTAMP(0)
PREV_SERVICE	VARCHAR(20)
CURR_SERVICE	VARCHAR(20)

Detailed Description

SUBJECT_ID, HADM_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient and HADM_ID is unique to a patient hospital stay.

TRANSFERTIME

TRANSFERTIME is the time at which the patient moved from the PREV_SERVICE (if present) to the CURR_SERVICE.

PREV_SERVICE, CURR_SERVICE

PREV_SERVICE and CURR_SERVICE are the previous and current service that the patient resides under.

TRANSFERS

Overview

Table source: Hospital database.

Table purpose: Physical locations for patients throughout their hospital stay.

Number of rows: 261,897

Links to:

- PATIENTS on SUBJECT_ID
- ADMISSIONS on HADM_ID
- ICUSTAYS on ICUSTAY_ID

Important considerations

- The ICUSTAYS table is derived from this table.
- Care units are defined based off the WARDID being associated with an ICU cost center.
- ICUs in the Beth Israel have moved throughout the years, and consequently the same WARDID may be considered as an ICU for patient A but not an ICU for patient B.

Table columns

Name	Postgres data type
ROW_ID	INT
SUBJECT_ID	INT
HADM_ID	INT
ICUSTAY_ID	INT

Name	Postgres data type
DBSOURCE	VARCHAR(20)
EVENTTYPE	VARCHAR(20)
PREV_CAREUNIT	VARCHAR(20)
CURR_CAREUNIT	VARCHAR(20)
PREV_WARDID	SMALLINT
CURR_WARDID	SMALLINT
INTIME	TIMESTAMP(0)
OUTTIME	TIMESTAMP(0)
LOS	INT

Detailed Description

SUBJECT_ID, HADM_ID, ICUSTAY_ID

Identifiers which specify the patient: SUBJECT_ID is unique to a patient, HADM_ID is unique to a patient hospital stay and ICUSTAY_ID is unique to a patient ICU stay.

DBSOURCE

DBSOURCE contains the original ICU database the data was sourced from. Patients admitted between 2001 - 2008 had their data managed by the CareVue information system, represented in this column as ‘carevue’. Patients admitted between 2008 - 2012 had their data managed by the Metavision system, represented in this column as ‘metavision’. Knowing the database source is important as the data archiving for these two databases differs in some cases.

EVENTTYPE

EVENTTYPE describes what transfer event occurred: ‘admit’ for an admission, ‘transfer’ for an inter-hospital transfer and ‘discharge’ for a discharge from the hospital.

PREV_CAREUNIT, CURR_CAREUNIT

PREV_CAREUNIT contains the care unit in which the patient previously resided. CURR_CAREUNIT contains the care unit in which the patient currently resides. The care unit is defined based upon the ward: if the ward is an ICU cost center, then the care unit defines the type of ICU. If the ward is *not* an ICU then in most cases the care unit is null. There are one or two exceptions to this rule. For example, NWARD is a ward for newborns.

The INTIME and OUTTIME of the transfer event correspond to the CURR_CAREUNIT. The PREV_CAREUNIT for each row is provided for convenience, and is identical to the CURR_CAREUNIT of the previous row (assuming the event is not an admission).

Care units include the following:

Care unit	**Description**
CCU	Coronary care unit
CSRU	Cardiac surgery recovery unit
MICU	Medical intensive care unit
NICU	Neonatal intensive care unit
NWARD	Neonatal ward
SICU	Surgical intensive care unit
TSICU	Trauma/surgical intensive care unit

PREV_WARDID, CURR_WARDID

PREV_WARDID and CURR_WARDID contain the previous and current ward in which the patient stayed. Note that the grouping of physical locations in the hospital database is referred to as a ward. Though in practice ICUs are not referred to as wards, the hospital database technically tracks ICUs as “wards with an ICU cost center". As a result, each ICU is associated with a WARDID, but not every WARDID is an ICU.

INTIME, OUTTIME

INTIME provides the date and time the patient was transferred into the current care unit from the previous care unit. OUTTIME provides the date and time the patient was transferred out of the current care unit.

LOS

LOS is the length of stay for the patient for the given ward stay, which may be within or outside of the ICU.

Patient Identifiers

Types of data in the database

Data within MIMIC were recorded during routine clinical care and *not* explicitly for the purpose of retrospective data analysis. This is a key point to keep in mind when analyzing the data.

There are two types of data in the database: static data and dynamic data. Static data is recorded once for a given identifier. An example of static data is the DOB column in the PATIENTS table. Each patient has only one date of birth, which does not change over time and is not recorded with an associated timestamp. An example of dynamic data is a patient's blood pressure, which is periodically measured during a hospital stay.

This distinction between static data and dynamic data is merely a helpful conceptual construct: there is *no* strict technical distinction between date of birth and heart rate. However, static data tends to not have an associated ITEMID (as there is no need to repeatedly record values for static data), whereas dynamic data have an ITEMID to facilitate efficient storage of repeated measurements.

Static data

SUBJECT_ID

The following is a list of static data available in the database for SUBJECT_ID in the PATIENTS table:

The only static data available for patients are their gender (GENDER), date of birth (DOB) and various dates of death (DOD, DOD_HOSP, DOD_SSN). These columns all occur in the PATIENTS table.

HADM_ID

The following is a list of static data available in the database for HADM_ID in the ADMISSIONS table:

- Admission time
- Discharge time
- Death time
- Admission type

ICUSTAY_ID

The following is a list of static data available in the database for ICUSTAY_ID in the ICUSTAYS table:

- INTIME - Time entered the ICU
- OUTTIME - Time left the ICU
- First care unit
- Last care unit

Hospital acquired data

The following tables were sourced from the hospital database, and contain information recorded in the hospital, but not necessarily during the patient's ICU stay:

- ADMISSIONS
- CALLOUT
- CPTEVENTS
- DIAGNOSES_ICD
- DRGCODES
- ICUSTAYS
- LABEVENTS
- MICROBIOLOGYEVENTS
- PATIENTS
- PRESCRIPTIONS
- PROCEDURES_ICD
- SERVICES
- TRANSFERS

ICU acquired data

The following tables were sourced from the ICU databases, and contain information only during a patient's ICU stay:

- DATETIMEEVENTS
- INPUTEVENTS_CV
- INPUTEVENTS_MV
- NOTEEVENTS
- OUTPUTEVENTS
- PROCEDUREEVENTS_MV

Externally acquired data

The DOD_SSN (which also contributes to the DOD column) is acquired from the social security death registry. It contains dates of death up to 90 days in the future for Metavision patients. It contains dates of death up to 4 years in the future for CareVue patients.

Data Sources

Patient level data: SUBJECT_ID

The following is a list of static data available in the database for SUBJECT_ID in the PATIENTS table:

- GENDER
- DOB
- DOD
- DOD_HOSP
- DOD_SSN

This list contains all the static data available for a single patient. Note that the DOD_SSN (which also contributes to the DOD column) is acquired from the social security death registry, i.e. an external source. It contains dates of death up to 90 days in the future for Metavision patients. It contains dates of death up to 4 years in the future for CareVue patients.

Hospital level data: HADM_ID

The following tables were sourced from the hospital database, and contain information recorded in the hospital, but not necessarily during the patient's ICU stay:

- ADMISSIONS
- CALLOUT
- CPTEVENTS
- DIAGNOSES_ICD
- DRGCODES
- ICUSTAYS
- LABEVENTS
- MICROBIOLOGYEVENTS
- PATIENTS

- PRESCRIPTIONS
- PROCEDURES_ICD
- SERVICES
- TRANSFERS

The following is a list of static data available in the database for HADM_ID in the ADMISSIONS table. This data is constant for a single hospital admission.

- ADMITTIME - The hospital admission time
- DISCHTIME - The hospital discharge time
- DEATHTIME - The date of death of the patient if they died within the hospital
- ADMISSION_TYPE - The type of admission: ELECTIVE, EMERGENCY, NEWBORN or URGENT (note that the NEWBORN value here does not perfectly identify newborns)
- ADMISSION_LOCATION - The location of the patient prior to hospital admission
- DISCHARGE_LOCATION - The location of the patient after hospital discharge
- INSURANCE - The patient's type of medical insurance
- LANGUAGE - The patient's primary language
- RELIGION - The patient's stated religion
- MARITAL_STATUS - The patient's marital status
- ETHNICITY - The patient's stated ethnicity
- DIAGNOSIS - A short description of the reason for the patient's admission

ICU level data: ICUSTAY_ID

The following tables were sourced from the ICU databases, and contain information only during a patient's ICU stay:

- DATETIMEEVENTS
- INPUTEVENTS_CV
- INPUTEVENTS_MV
- NOTEEVENTS
- OUTPUTEVENTS

- PROCEDUREEVENTS_MV

The following is a list of static data available in the database for ICUSTAY_ID in the ICUSTAYS table:

- DBSOURCE - The ICU database from which the patient exists in
- FIRST_CAREUNIT - The first unit that cared for the patient (all ICUs except NWARD)
- LAST_CAREUNIT - The last care unit that cared for the patient (all ICUs except NWARD)
- FIRST_WARDID - An integer representing the first physical location of the patient
- LAST_WARDID - An integer representing the last physical location of the patient
- INTIME - Time entered the ICU
- OUTTIME - Time left the ICU
- LOS - The patient's ICU length of stay

Times

Time types

Time in the database is stored with one of two suffixes: TIME and DATE. If a column has TIME as the suffix, e.g. CHARTTIME, then the data resolution is down to the minute. If the column has DATE as the suffix, e.g. CHARTDATE, then the data resolution is down to the day. That means that measurements in a CHARTDATE column will always have 00:00:00 has the hour, minute, and second values. This does *not* mean it was recorded at midnight: it indicates that we do not have the exact time, only the date.

Date shifting

All dates in the database have been shifted to protect patient confidentiality. Dates will be internally consistent for the same patient, but randomly distributed in the future. Dates of birth which occur in the present time are *not* true dates of birth. Furthermore, dates of birth which occur before the year 1900 occur if the patient is older than 89. In these cases, the patient's age at their first admission has been fixed to 300.

Time columns in the database

CHARTTIME vs STORETIME

Most data, with the exception of patient related demographics, are recorded with a time indicating when the observation was made: CHARTTIME. CHARTTIME dates back to the use of paper charts: in order to facilitate efficient observations by nursing staff, the day was separated into hourly blocks, and observations were recorded within these hourly blocks. Thus, any time one performed a measurement between the hours of 04:00 and 05:00, the data would be charted in the 04:00 block, and so on. This concept has carried forward into the electronic recording of data: even if data is recorded at 04:23, in many cases it is still charted as occurring at 04:00.

STORETIME provides information on the recording of the data element itself. All observations in the database must be validated before they are archived into the patient medical record. The STORETIME provides the exact time that this validation occurred. For example, a heart rate may be charted at 04:00, but only validated at 04:40. This indicates that the care provider validated the measurement at 4:40 and indicated that it was a valid observation of the patient at 04:00. Conversely, it's also possible that the STORETIME occurs *before* the CHARTTIME. While a Glasgow Coma Scale may be charted at a CHARTTIME of 04:00, the observation may have been made and validated slightly before (e.g. 3:50). Again, the validation implies that the care staff believed the measurement to be an accurate reflection of the patient status at the given CHARTTIME.

Summing up: CHARTTIME vs. STORETIME

CHARTTIME is the time at which a measurement is *charted*. In almost all cases, this is the time which best matches the time of actual measurement. In the case of continuous vital signs (heart rate, respiratory rate, invasive blood pressure, non-invasive blood pressure, oxygen saturation), the CHARTTIME is usually exactly the time of measurement. STORETIME is the time at which the data is recorded in the database: logically it occurs after CHARTTIME, often by hours, but usually not more than that.

CHARTDATE

CHARTDATE is equivalent to CHARTTIME, except it does not contain any information on the time (all hour, minute, and seconds are 0 for these measurements).

ADMITTIME, DISCHTIME, DEATHTIME

ADMITTIME and DISCHTIME are the hospital admission and discharge times, respectively. DEATHTIME is the time of death of a patient if they died *in* hospital. If the patient did not die within the hospital for the given hospital admission, DEATHTIME will be null.

CREATETIME, UPDATETIME, ACKNOWLEDGETIME, OUTCOMETIME, FIRSTRESERVATIONTIME, CURRENTRESERVATIONTIME

CREATETIME is the time at which an ICU discharge was requested for a given patient. UPDATETIME is the time which the ICU discharge request was updated. ACKNOWLEDGETIME was the time at which the discharge request was acknowledged by the transfers team. OUTCOMETIME is the time at which the ICU discharge request was completed (with an outcome of ‘Discharged’ or ‘Canceled’).

FIRSTRESERVATIONTIME and CURRENTRESERVATIONTIME only occur for patients who require certain locations in the hospital.

INTIME, OUTTIME

INTIME and OUTTIME provide the time at which a patient entered and exited the given unit. In the ICUSTAYS table, the unit is always an ICU. In the TRANSFERS table, the unit can be any ward in the hospital.

STARTTIME, ENDTIME

For events which occur over a period of time, STARTTIME and ENDTIME provide the beginning and end time of the event. For medical infusions, these columns indicate the period over which the substance was administered.

COMMENTS_DATE

COMMENTS_DATE provides the time at which a cancel or edit comment was made for a given order.

DOB, DOD, DOD_HOSP, DOD_SSN

DOB is the patient's date of birth. If the patient is older than 89, their date of birth is set to 300 at their first admission. DOD is the patient's date of death: sourced either from the hospital database (DOD_HOSP) or the social security database (DOD_SSN).

TRANSFERTIME

TRANSFERTIME is the time at which the patient's service changes.

Inputs and outputs

Inputs and outputs are extremely useful when studying intensive care unit patients. Inputs are any fluids which have been administered to the patient: such as oral or tube feedings or intravenous solutions containing medications. Outputs are fluids which have either been excreted by the patient, such as urine output, or extracted from the patient, for example through a drain. These data were the most complicated to handle technically in the MIMIC-III data.

The MIMIC-III database contains information from two distinct critical care information systems: Philips CareVue and iMDSoft Metavision. These two databases store data in different ways. In the descriptions in this document, data will be referred to as being sourced either from “CareVue” or “Metavision” to differentiate between the different systems.

INPUTEVENTS_CV, INPUTEVENTS_MV, OUTPUTEVENTS

Inputs exist in two separate tables: INPUTEVENTS_CV and INPUTEVENTS_MV. INPUTEVENTS_CV contains CareVue inputs, while INPUTEVENTS_MV contains Metavision inputs. Results from these tables can be unioned as the same patient stay never occurs in both tables. Concretely, a unique ICUSTAY_ID only occurs in *one* out of the two tables. However, if a patient has two ICU stays, one in CareVue and one in Metavision, then they could appear in both tables. As a result, a unique SUBJECT_ID could occur in both tables. All outputs for both Metavision and CareVue patients have been merged into the OUTPUTEVENTS table.

Outputs

Outputs are recorded in a reasonably consistent manner in both the CareVue and Metavision data. The time at which the output is measured is recorded in the CHARTTIME column. There is no start time recorded with outputs - CHARTTIME corresponds to the time that the volume had been output by. The volume of output is recorded in the VALUE column, and the unit of measurement is provided in the VALUEUOM column (usually milliliters, or mL). It is usually reasonable to assume that any output recorded is for the interval between the current CHARTTIME and the previous CHARTTIME for the same item.

Inputs

Inputs are handled differently by CareVue and Metavision. For CareVue data, only the CHARTTIME is available. Second, the RATE and AMOUNT columns are *asynchronous*, and originally stored in different tables. Volumes of input (e.g. 50 mL of normal saline) would be stored in one table (the original IOEVENTS), and would usually be recorded every hour (though sometimes the period was longer). Conversely, the RATE of the drug would be stored separately (in MEDEVENTS), and only updated when a change or verification of the rate was made by clinical staff. As a result, the raw data looked something similar to:

CHARTTIME	VOLUME	VOLUMEUOM	CHARTTIME	RATE	RATEUOM
			09:00	1	mL/min
10:00	60	mL			
11:00	60	mL			
			11:30	0.5	mL/min
12:00	45	mL			

Here, the volume is recorded only every hour, and no start time is available. However, it's reasonable to assume that the volume measurement corresponds to an hour. Next, we can see that the rate was titrated to 0.5, and for the period between 11:00 to 12:00 there was half an hour of delivery at 1 mL/min, and half an hour of delivery at 0.5 mL/min, resulting in a total volume of 45 mL delivered for the past hour.

Summing up, for CareVue data, the rate and volume will be asynchronous, and only the CHARTTIME will be available. For rates, the CHARTTIME will correspond to a start time (when the drug was set to that rate). For volumes, the CHARTTIME will correspond to an end time.

For Metavision data, there is no concept of a volume in the database: only a RATE. All inputs are recorded with a STARTTIME and an ENDTIME. As a result, the volumes in the database for Metavision patients are *derived* from the rates. Furthermore, exact start and stop times for the drugs are easily deducible.

Details of the merging process

The difficulty in merging the databases arose due primarily to two factors: the lower resolution of information archiving in the CareVue system, and the different definition of an 'order' in the databases.

The aim of this section is to provide all the detail into how these data were merged: this information is not necessary to understand for the purposes of using the database, but will provide insight into the format of the IOEVENTS table. We welcome suggestions from the community on improving the format and usability of the table.

Philips CareVue

The CareVue system stored input/output (IO) data across five tables: IOEVENTS, MEDEVENTS, ADDITIVES, SOLUTIONS and DELIVERIES. Each time a new order for a drug was recorded in the database, the ADDITIVES, SOLUTIONS and DELIVERIES tables would be populated with information regarding the order. The data archival format is best described with an example.

Patient A has been recently admitted to the ICU and is to be administered noradrenaline to restore their blood pressure to a value of at least 60 mmHg. The route of administration is intravenous, i.e. the drug is to be pumped into the patient's blood stream directly through a line inserted in a vein. The nurse would prepare a solution of 250 mL sodium chloride (NaCl) to contain the drug. The SOLUTIONS table would consequently contain an entry of 250 mL NaCl for patient A at the time the nurse prepared the solution. The nurse would then mix in the drug into the solution: in this example let's say 8 mg of noradrenaline. The 8 mg of noradrenaline would be recorded in the ADDITIVES table, and an identifier would be recorded which linked the 8 mg of noradrenaline to the 250 mL solution of NaCl. Finally, the nurse would optionally set an initial delivery rate and route: in this case it could be 10 mL/hr intravenously. These would populate the RATE and ROUTE columns in the DELIVERIES table. Note that DELIVERIES data is less frequently present and was not consistently recorded.

Now that the solution has been prepared, the nurse can begin administering the drug to the patient. The time is now 18:20. An initial entry of 0 mL is recorded in the IOEVENTS table at 18:00 (this usually occurs, but it has not been verified that this *always* occurs). The nurse begins at a rate of 1 mcg/kg/min. The MEDEVENTS contains an entry at a CHARTTIME of 18:20 for a rate of 1 mcg/kg/min. Five minutes later, at 18:25, the nurse notes that the blood pressure is still lower than the desired 60 mmHg and increases the dose. The nurse raises the dose to 2 mcg/kg/min, and MEDEVENTS records the new dose of 2 mcg/kg/min at 18:25. The nurse checks again and notes that the blood pressure has reached the target value of 60 mmHg and ceases titration of the drug. At 19:00, the volume of drug administered to the patient is recorded. If the patient weighed 100 kg, then the amount of drug administered would be equal to:

5 min * 1 mcg/kg/min * 100 kg + 25 min * 2 mcg/kg/min * 100 kg = 500 mcg + 5000 mcg = 5500 mcg = 5.5 mg

The amount of solution administered would be equal to:

(5 min * 1 mcg/kg/min * 100 kg + 25 min * 2 mcg/kg/min * 100 kg) * 250 mL / 8 mg = 5.5 mg * 250 mL / 8 mg = 171.875 mL

Consequently, IOEVENTS would record 171.875 mL at 19:00. If the patient continued at the same rate for this drug, then IOEVENTS would record (60 * 2 * 100 * 250 / 8000) = 375 mL at 20:00. Note that:

- Unless the rate is updated, no new entry in MEDEVENTS exists
- IOEVENTS always records a value on an hourly basis, regardless of how long the IO event has been present
- 375 mL is larger than the original 250 mL bag: somewhere during administration the nurse would need to replace the empty bag with a new solution of the same formulation

Note that the changing of the bag is sometimes, but not always, recorded in the ADDITIVES/SOLUTIONS/DELIVERIES table as a new order. It usually occurs if the drug has been discontinued, and then re-prescribed later in the patient’s stay.

Metavision

Metavision records IO data using two tables: RANGESIGNALS and ORDERENTRY. These tables do not appear in MIMIC-III as they have been merged to form the INPUTEVENTS_MV table. RANGESIGNALS contains recorded data elements which last for a fixed period of time. Furthermore, the RANGESIGNALS table recorded information for each component of the drug separately. For example, for a noradrenaline administration there would be two components: a main order component (noradrenaline) and a solution component

(NaCl). The STARTTIME and ENDTIME of RANGESIGNALS indicated when the drug started and finished. *Any* change in the drug rate would result in the current infusion ending, and a new STARTTIME being created.

Let's return to our example above of the patient being given noreadrenaline. The STARTTIME for the solution (NaCl) and the drug (noradrenaline) would be 18:20. The rate of the drug would be listed as 1 mcg/kg/min, and the rate of the solution would be listed as 10 mL/hr. The nurse, as before, decides to increase the drug rate at 18:25 to 2 mcg/kg/min. As a result, the ENDTIME for the two rows corresponding to the solution (NaCl and noreadrenaline) is set to 18:25. Two new rows are generated with a STARTTIME of 18:25. These two new rows would continue until either (i) the drug rate was changed or (ii) the drug was delivery was discontinued. The ORDERID column would be identical for each instantiation of NaCl and noradrenaline which corresponded to the same solution/rate. That is, for the infusion given between 18:20 and 18:25, both NaCl and noreadrenaline would have the same ORDERID. LINKORDERID would further link the drug across all administrations, even when the rate is changed. The following table demonstrates these concepts:

Item	**STARTTIME**	**ENDTIME**	**RATE**	**RATEUOM**	**ORDERID**	**LINKORDERID**
NaCl	18:20	18:25	1	mcg/kg/min	8003	8003
Noradrenaline	18:20	18:25	10	ml/hr	8003	8003
NaCl	18:25	20:00	2	mcg/kg/min	8020	8003
Noradrenaline	18:25	20:00	20	ml/hr	8020	8003

Note that ORDERID links items occurring at the same time which correspond to the same solution, while LINKORDERID links all these solutions together across time. Note also that LINKORDERID is equal to the first ORDERID which occurs for the solution, as above.

Glossary of terms

Term	Description
BIDMC	Beth Israel Deaconess Medical Center
CCU	Coronary Care Unit
CVICU	Cardiovascular Intensive Care Unit
EHR	Electronic Health Record
ICU	Intensive Care Unit
LCP	Laboratory for Computational Physiology
MICU	Medical Intensive Care Unit
MIMIC	Medical Information Mart for Intensive Care (*previously Multiparameter Intelligent Monitoring in Intensive Care*)
NICU	Neonatal Intensive Care Unit
SICU	Surgical Intensive Care Unit

References

[1] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635

[2] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).