Introduction to MIMIC-3 Database

Jeong-Min Lee

jlee@cs.pitt.edu

2016. 12. 7.

Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Introduction

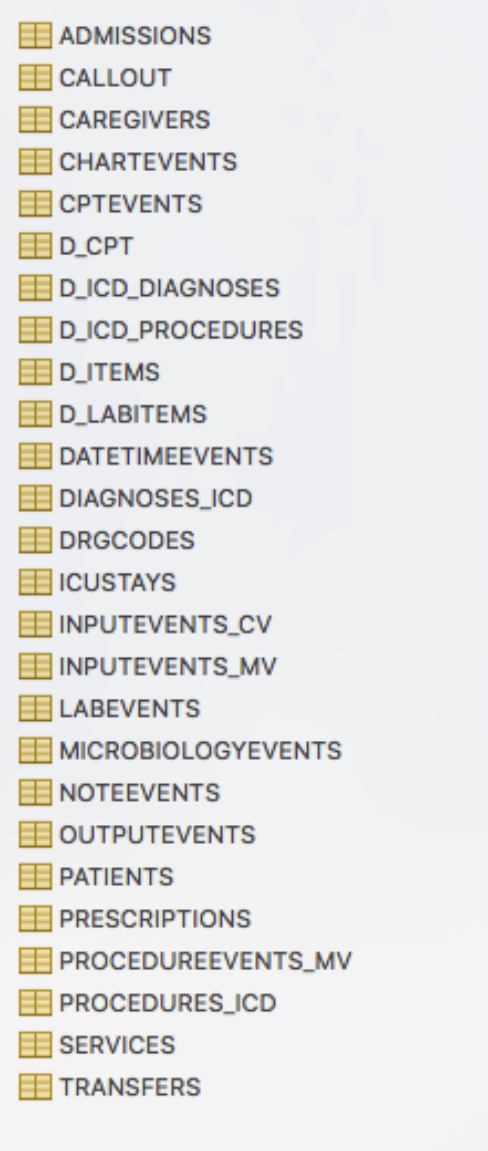
- MIMIC-3: Medical Information Mart for Intensive Care 3
- Publicly available deidentified patient record.
- Records on 46520 patient stayed in critical care units of Beth Israel Deaconess Medical Center (at Boston) between 2001 and 2012.
- It includes demographics, vital sign measurements, lab test results, procedures, medications, and clinical notes.
- This slide is largely based on the information provided in MIMIC-3 version 1.4 tutorial website: http://mimic.physionet.org

Introduction

- There is a MIMIC ‘version 2’
  - A little bit different table structure.
  - But the version 2 comes with a clinical database and a waveform database.
  - Waveform database: many of records of continuously digitized physiologic waveforms and simultaneously recorded time series (trends) of physiologic measurements.
    - 3TB of data in all.
    - Signals (125 samples per second):
      - ECG (electrocardiographic) waveforms
      - BP (continuous blood pressure) waveforms
      - Raw output of fingertip plethysmograph
      - Respiration waveforms
    - Numerics/Trends (1 sample/second or 1 sample/minute):
      - BP (systolic, diastolic, and mean)
      - CO: cardiac output
      - CO2: carbon dioxide
      - HR: heart rate
      - Respiration rate
      - SpO2: oxygen saturation (from fingertip plethysmography)
      - ST: ECG ST segment levels
      - Temperature
- It will not be covered in this presentation

Overview

- The MIMIC-3 database is consisted of 26 tables.
- Each table contains each patient record (at each row) with specific field (columns)
- Tables starts with ‘D_’ are dictionaries and provide definitions for identifiers.
- “_MV” and “CV” in table names are representing different information systems used to collect data.
  - CV: Philips Carevue, 2001-2008
  - MV: iMDSoft Metavision, 2008-2012



Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Patient Information

- Each patient is unique with its own “subject_id”.
- Each hospital admission of a patient is unique with “hadm_id”.
- Each ICU stay of a patient is unique with “icustay_id”.
- That means,
  - One subject_id can be associated with multiple hadm_ids when a patient had multiple admissions.
  - One hadm_id can be linked to multiple icustay_id when a patient had a multiple ICU stays during an admission. (e.g., transferring between multiple ICUs)

SUBJECT_ID	HADM_ID	ICUSTAY_ID
22067	100566	258026
22067	100566	253296
22067	100566	232637

Patient Information

- PATIENTS table
  - Gender, DOB (date of birth)
  - DOD (date of death), expire flag (whether a patient died or not)
  - Age of certain patient of a point of time in the record can be calculated by subtracting a certain record time – dob.
  - Note that all dates in the database is shifted randomly with deidentification process. But it is consistent throughout a patient's records.

Patient Information

- ADMISSIONS table
  - Admit and discharge time, death time if died in the admission
  - Admission and discharge location in hospital
  - Insurance, language, ethnicity, and marital status
  - Diagnosis, ** but this diagnosis is usually assigned by the admitting clinician and **not** use a systematic ontology (such as ICD9 code)
  - Final diagnosis can be found in DIAGNOSIS_ICD table.

- ICUSTAYS table
  - In/Out time
  - LOS (length of stay), values are normalized to 1.0 = 24hours

Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Medication

- Three tables contain information on medication:
  - 1) PRESCRIPTIONS
  - 2) INPUTEVENTS_CV
  - 3) INPUTEVENTS_MV

Medication

- PRESCRIPTIONS table
  - Medication related order entries.
  - Drug names, NDC code (11digit version)
    - National Drug Code (NDC) is originally 10-digit,3 segment numeric identifier.
    - Centers for Medicare and Medicaid Services (CMS) has created 11-digit NDC derivative. This format has selected by HIPAA regulation thus other governmental agencies' database such as UMLS (which contains RXNORM, NDF-RT ontologies) are using this 11-digit NDC.

DRUG	DRUG_NAME_POE	DRUG_NAME_GENERIC	FORMULARY_DRUG_CD	GSN	NDC
Ipratropium Bromide MDI	Ipratropium Bromide MDI	Ipratropium Bromide HFA	IPRAPF	059081	00597008717
Albuterol Inhaler	Albuterol Inhaler	Albuterol Inhaler (ProAir)	PROAIR8.5H	028090	59310057920
Heparin	Heparin	Heparin Sodium	HEPA5I	006549	63323026201

Medication

- INPUTEVENTS tables
  - INPUTs are any fluids which have been administered to the patient
    - such as oral or tube feedings or intravenous solutions containing medications.
  - Inputs exist in two separate tables:
    - INPUTEVENTS_CV contains CareVue inputs
    - INPUTEVENTS_MV contains Metavision inputs

Medication

- INPUTEVENTS_CV table
  - For CareVue data, only the CHARTTIME is available.
  - The RATE and AMOUNT columns are *asynchronous*, and originally stored in different tables. (In MIMIC-3, these are in one table)
  - Volumes of input (e.g. 50 mL of normal saline) would usually be recorded every hour (though sometimes the period was longer).
  - RATE of the drug only updated when a change or verification of the rate was made by clinical staff.

Medication

- INPUTEVENTS_CV table (cont.)

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			

  - Assuming this is a patient's inputevent records of a certain medication.

Medication

- INPUTEVENTS_CV table (cont.)

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			

  - The volume is recorded only every hour
    - No start time is available. (CHARTTIME of volume is end time)

Medication

- INPUTEVENTS_CV table (cont.)

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			

  - The volume is recorded only every hour
    - No start time is available.
    - However, it's reasonable to assume that the volume measurement corresponds to an hour.

Medication

- INPUTEVENTS_CV table (cont.)

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			

- Next, we can see that the rate was titrated to 0.5
  - 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 an hour between 11:00 and 12:00.

Medication

- INPUTEVENTS_CV table (cont.)

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			

  - Remember that VOLUME would usually be recorded every hour.
  - RATE is only updated when a change or verification of the rate was made by clinical staff.

Medication

- INPUTEVENTS_CV table (cont.)

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			

  - 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

Medication

- INPUTEVENTS_CV table (cont.)
  - ORDERID links multiple items contained in the same solution together.
    - For example, when a drug of noradrenaline and a solution of NaCL is administered.
    - Both noradrenaline and NaCL occur on distinct rows but will have the same ORDERID.

CHARTTIME	ITEM	ORDERID
10:00	noradrenaline (drug)	201
10:00	NaCl (solution)	201

Medication

- INPUTEVENTS_CV table (cont.)
  - LINKORDERID links the same order across multiple instantiations
    - For example, when the rate of delivery for the solution with noradrenaline and NaCL is changed
    - Two new rows which share the same new ORDERID will be generated, but the LINKORDERID will be the same.

CHARTTIME	ITEM	ORDERID	LINKORDERID	RATE
10:00	noradrenaline	201	5001	1.0
10:00	NaCl	201	5001	1.0
11:00	noradrenaline	202	5001	0.5
11:00	NaCl	202	5001	0.5

Medication

- INPUTEVENTS_MV table
  - 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

Medication

- INPUTEVENTS_MV table (cont.)

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

- The STARTTIME for the solution (NaCl) and the drug (noradrenaline) would be 18:20.
- Rate for the drug would be 1 mcg/kg/min and 10 ml/hr for solution. (mcg: microgram)

Medication

- INPUTEVENTS_MV table (cont.)

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

- The nurse 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.

Medication

- INPUTEVENTS_MV table (cont.)

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

- 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.

Medication

- INPUTEVENTS_MV table (cont.)

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

- 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.

Medication

- INPUTEVENTS_MV table (cont.)

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

- LINKORDERID would further link the drug across all administrations, even when the rate is changed.

Topics

- Introduction
- Overview
- Patient Information
- Medications
- **Lab Tests**
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Lab tests

- Three tables contain lab test record
  - CHARTEVENTS
  - LABEVENTS
  - MICROBIOLOGYEVENTS
- Two tables contain itemid label information
  - D_ITEMS (for CHARTEVENTS and MICROBIOLOGYEVENTS)
  - D_LABITEMS (for LABEVENTS)

Lab tests

- CHARTEVENTS table
  - CHARTEVENTS table contains all the charted data available for patients.
  - It contains 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 more than lab tests.
  - Furthermore, even though laboratory values are captured elsewhere (LABEVENTS), they are frequently repeated within CHARTEVENTS.
  - In cases where there is disagreement between measurements, LABEVEVENTS table should be taken as the ground truth.

ITEMID	CHARTTIME	STORETIME	CGID	VALUE	VALUENUM	VALUEUOM
220051	2157-10-21 12:15:00	2157-10-21 12:15:00	16978	74	74	mmHg
226512	2157-10-21 12:15:00	2157-10-21 12:15:00	16978	66.8	66.8	kg
220046	2157-10-21 12:43:00	2157-10-21 12:43:00	16978	110	110	bpm

Lab tests

- CHARTEVENTS table (cont.)
  - Itemid: identifier for a measurement type in the database
    - The label and definition of itemid can be found in D_ITEMS table
    - For example, itemid 212 in CHARTEVENTS is ‘heart failure’
  - Value: the value measured for the item.
    - If this value is numeric, then VALUENUM contains the same data in a numeric format.

Lab tests

- LABEVENTS
  - This table contains information regarding laboratory based measurements.
  - Label of each itemid can be found in ‘D_LABITEMS’ table.
  - The data gathering process is as follows
    - 1) Clinical staff acquires a fluid form a site in the patient body
    - 2) The fluid is barcoded to associated with the patient and timestamped
    - 3) The lab analysis is run and result is returned within 4-12 hours.
  - The charttime column records when an observation is created.
    - It is closest proxy to the time the data was actually measured.

HADM_ID	ITEMID	CHARTTIME	VALUE	VALUENUM	VALUEUOM	FLAG
NULL	51006	2101-10-04 15:24:00	42	42	mg/dL	abnormal
NULL	51137	2101-10-04 15:24:00	NORMAL	NULL	NULL	NULL
NULL	51146	2101-10-04 15:24:00	0.5	0.5	%	NULL

Lab tests

- LABEVENTS (cont.)
  - Value field contains the value measure for the lab item.
  - FLAG field indicated whether the value is considered abnormal or not using pre-defined thresholds.
  - A patient without hospital admission ID (hadm_id) is a lab value obtained as outpatient.

Lab tests

- D_LABITEMS
  - Contains identifiers associated with lab measurements.
  - Each entry has LOINC codes.

  - Fluid: the substance which the measurement was made.
  - Category: higher level info as to the type of measurement.

	ITEMID	LABEL	FLUID	CATEGORY	LOINC_CODE
1	50800	SPECIMEN TYPE	BLOOD	BLOOD GAS	NULL
2	50801	Alveolar-arterial Gradient	Blood	Blood Gas	19991-9
3	50802	Base Excess	Blood	Blood Gas	11555-0

Lab tests

- MICROBIOLOGYEVETNS
  - Contains microbiology information, including tests performed and sensitivities.
  - CHARTTIME: time at which an observation was charted, and is usually the closest proxy to the time the data was actually measured.
  - SPEC_ITEMID and SPEC_TYPE_DESC : info on specimen
  - ORG_ITEMID and ORG_NAME: info on organism
  - AB_ITEMID and AB_NAME: info on antibody
  - INTERPRETATION: "S" is sensitive, "R" is resistant, "I" is intermediate, and "P" is pending

CHARTTIME	SPEC_ITEMID	SPEC_TYPE_DESC	ORG_ITEMID	ORG_NAME	ISOLATE_NUM	AB_ITEMID	AB_NAME
2165-06-03 08:28:00	70079	URINE	80004	KLEBSIELLA PNEUMONIAE	1	90022	AMPICILLIN/SULBACTAM
2165-06-03 08:28:00	70079	URINE	80004	KLEBSIELLA PNEUMONIAE	1	90019	CIPROFLOXACIN
2165-06-03 08:28:00	70079	URINE	80004	KLEBSIELLA PNEUMONIAE	1	90018	CEFTRIAXONE
2165-06-03 08:28:00	70079	URINE	80004	KLEBSIELLA PNEUMONIAE	1	90017	CEFTAZIDIME
2165-06-03 08:28:00	70079	URINE	80004	KLEBSIELLA PNEUMONIAE	1	90013	TOBRAMYCIN

Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Outputs

- 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.
- All outputs records are in OUTPUTEVENTS table.
  - 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.

Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Diagnosis

- DIAGNOSES_ICD table contains tuples of patient id (subject_id), admission id (hadm_id), and ICD-9 code.
- No timestamp information is available.
- 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.
- Sequence number: ICD diagnoses are ordered by priority - and the order does have an impact on the reimbursement for treatment.
- D_ICD_DIAGNOSES table contains label for each ICD-9 code.

Procedures

- PRECEDURES_ICD table contains procedure record.
  - Similar structure to DIAGNOSES_ICD table:
    - Patient id, admission id, ICD-9 procedure code, and sequence number.
  - D_ICD_PROCEDURES table contains label for ICD-9 procedure codes.
  - 3882 different kinds of procedures are used in this table.

- PROCEDUREEVENTS_MV table also contains procedure records.
  - Patient procedures for the subset of patients who were monitored in the ICU using the iMDSoft MetaVision system.
  - 125 different kinds of procedures are used in this table.

Procedures

- CPTEVEENTS table contains current procedural terminology (CPT) codes which facilitate billing purpose.
  - CPT_CD : original CPT code
  - CPT_NUMBER: numeric version of CPT_CD
  - SECTIONHEADER: a category of the CPT code.
    - D_CPT table contains label for the CPT code.

Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Clinical Notes

- NOTEEVENTS table contains clinical free text note.
- Including nursing and physician notes, ECG reports, radiology reports, and discharge summaries.
- Full category:
  - Case Management
  - Consult
  - Discharge summary
  - ECG
  - Echo (echoencephalogram)
  - General
  - Nursing
  - Nursing/other
  - Nutrition
  - Pharmacy
  - Physician
  - Radiology
  - Rehab Services
  - Respiratory
  - Social Work

Other tables

- There are some tables that I didn't mention in this slides.
- A brief is:
  - CALLOUT: Information regarding when a patient was cleared for ICU discharge and when the patient was actually discharged.
  - CAREGIVERS: Every caregiver who has recorded data in the database (defines CGID).
  - DATETIMEEVENTS: All recorded observations which are dates, for example time of dialysis or insertion of lines.
  - DRGCODES: Diagnosis Related Groups (DRG), which are used by the hospital for billing purposes.
  - SERVICES: The clinical service under which a patient is registered.
  - TRANSFERS: Patient movement from bed to bed within the hospital, including ICU admission and discharge.
- (The full brief is from http://www.nature.com/articles/sdata201635/tables/4)

Topics

- Introduction
- Overview
- Patient Information
- Medications
- Lab Tests
- Outputs
- Diagnosis and Procedures
- Clinical Notes
- How to Get Access?

Getting Access

- 1. Complete the required training course
  - Complete CITI “Data or Specimens Only Research” course
    - https://www.citiprogram.org/index.cfm?pageID=154&icat=0&ac=0
    - In the Human Subjects training category, select the “Data or Specimens Only Research” course
    - Complete the course and save a copy of your certificate.
  - Complete Pitt Required Educational modules for working with Biomedical data.
    - To be announced by Prof. Milos Hauskrecht
- 2. Request access to MIMIC-III
  - Create an account on PhysioNet using the following link: https://physionet.org/pnw/login
  - Follow the instructions on PhysioNet to apply for access to the MIMIC-III project, remembering to provide your CITI training certificate: https://physionet.org/works/MIMICIIIClinicalDatabase/access.shtml
  - When your application has been approved you will receive emails containing instructions for downloading the database from PhysioNetWorks.
  - Approval may take several weeks.

Questions?