MIMIC-III
A Freely Available
Critical Care Database

Tom J. Pollard
Jesse Raffa

MIT Laboratory for Computational Physiology,
Institute for Medical Engineering and Science
Research opportunity

- Huge volumes of data are captured daily

- …data that could be used to discover new knowledge for the benefit of patients
but, this data is inaccessible to researchers
MIMIC is an openly available dataset developed by the MIT Lab for Computational Physiology, comprising deidentified health data associated with ~40,000 critical care patients. It includes demographics, vital signs, laboratory tests, medications, and more.
MIMIC-III

- 2001 - 2012
- Waived consent for data collection
- ~40,000 patients
- Data extracted from digital systems
<table>
<thead>
<tr>
<th>Code status</th>
<th>Full code</th>
<th>Comfort measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCS: Verbal</td>
<td>Oriented</td>
<td>Incomprehensible sounds</td>
</tr>
<tr>
<td>GCS: Moto</td>
<td>Obeys commands</td>
<td>Flex-withdraws</td>
</tr>
<tr>
<td>GCS: Eye</td>
<td>Spontaneously</td>
<td>None</td>
</tr>
<tr>
<td>Platelet, K/uL</td>
<td>48, 53</td>
<td>Oriented</td>
</tr>
<tr>
<td>Creatinine, mg/dL</td>
<td>0.7, 0.7</td>
<td>Obeys commands</td>
</tr>
<tr>
<td>White blood cell, K/uL</td>
<td>9.1, 12.4</td>
<td>Spontaneously</td>
</tr>
<tr>
<td>Neutrophil, %</td>
<td>37, 37</td>
<td>Confused</td>
</tr>
<tr>
<td>GCS: Verbal</td>
<td>46</td>
<td>Obeys commands</td>
</tr>
<tr>
<td>GCS: Moto</td>
<td>0.7</td>
<td>To speech</td>
</tr>
<tr>
<td>GCS: Eye</td>
<td>16.8</td>
<td>Confused</td>
</tr>
<tr>
<td>Platelet, K/uL</td>
<td>45, 0.8</td>
<td>Oriented</td>
</tr>
<tr>
<td>Creatinine, mg/dL</td>
<td>0.8, 23.2</td>
<td>Obeys commands</td>
</tr>
<tr>
<td>White blood cell, K/uL</td>
<td>9.1, 12.4</td>
<td>Spontaneously</td>
</tr>
<tr>
<td>Neutrophil, %</td>
<td>37, 37</td>
<td>Confused</td>
</tr>
<tr>
<td>Morphine Sulfate</td>
<td>10.0mL/hour</td>
<td>Oriented</td>
</tr>
<tr>
<td>Vancomycin (1 dose)</td>
<td>1mg/min</td>
<td>Obeys commands</td>
</tr>
<tr>
<td>Piperacillin (1 dose)</td>
<td>0.5mg/min</td>
<td>To speech</td>
</tr>
<tr>
<td>NaCl 0.9%</td>
<td>50mL/hour</td>
<td>Confused</td>
</tr>
<tr>
<td>Amiodarone</td>
<td>25mL/hour</td>
<td>Oriented</td>
</tr>
<tr>
<td>Dextrose 5%</td>
<td>25mL/hour</td>
<td>Obeys commands</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measurement, absolute value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart rate</td>
</tr>
<tr>
<td>O2 saturation</td>
</tr>
<tr>
<td>NIBP, mean</td>
</tr>
<tr>
<td>Respiratory rate</td>
</tr>
<tr>
<td>Intake volume, dL</td>
</tr>
<tr>
<td>Output volume, dL</td>
</tr>
</tbody>
</table>

Time after admission to the intensive care unit, hours

GCS: Verbal
GCS: Moto
GCS: Eye
Platelet, K/uL
Creatinine, mg/dL
White blood cell, K/uL
Neutrophil, %
Morphine Sulfate
Vancomycin (1 dose)
Piperacillin (1 dose)
NaCl 0.9%
Amiodarone
Dextrose 5%
Admission Date: [**2952-11-3**]
Discharge Date: [**2952-11-9**]

Date of Birth: [**2887-7-23**]  
Sex: F

Service: MEDICINE

Allergies:
No Known Allergies / Adverse Drug Reactions

Attending: [**First Name3 (LF) 3925**]

Chief Complaint:
Sepsis, respiratory distress

Major Surgical or Invasive Procedure:
None

History of Present Illness:
F w/ h/o metastatic breast cancer to breast and lungs currently receiving CMT, brought to the ED by rehab for abnormal labs. She was found to be neutropenic, anemia and thrombocytopenic. At the rehab, vitals were reportedly T 100.4, HR 107, BP 92/42. There is also a concern for possible...
Accessing MIMIC

Two key steps to gaining access to MIMIC:

• **complete a online course in protecting human research participants** that covers Health Insurance Portability and Accountability Act (HIPAA) requirements

• **sign a data use agreement**, which outlines appropriate data usage and security standards, and forbids efforts to identify individual patients.
Relational database
(a collection of linked spreadsheets)

Each patient has a unique subject_id. This is the primary key (pk).

patients
subject_id (pk)
gender
dob
dod

admissions
subject_id (fk)
hadm_id (pk)
intime
outtime

subject_id is a foreign key (fk) here. It references a primary key in the patients table.
Patient tracking tables

- **patients**
  - subject_id

- **admissions**
  - hadm_id

- **icustays**
  - icustay_id
  - icustay_id
  - icustay_id
  - icustay_id
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>chartevents</td>
<td>Charted observations for a patient</td>
</tr>
<tr>
<td>labevents</td>
<td>Lab measurements both within hospital and outpatient clinics</td>
</tr>
<tr>
<td>inputevents</td>
<td>Input fluids (e.g. intravenous medications)</td>
</tr>
<tr>
<td>microbiology events</td>
<td>Microbiology measurements and sensitivities</td>
</tr>
<tr>
<td>noteevents</td>
<td>Deidentified patient notes</td>
</tr>
</tbody>
</table>
Other data tables

- **diagnoses_icd**: Hospital assigned diagnosis codes
- **procedures_icd**: Hospital assigned procedure codes
- **caregivers**: Caregivers who have recorded data
- **prescriptions**: Medications ordered for a patient
Reusable code

```sql
DROP MATERIALIZED VIEW IF EXISTS angus_sepsis CASCADE;
CREATE MATERIALIZED VIEW angus_sepsis as
-- ICD-9 codes for infection - as sourced from Appendix 1 of above paper
WITH infection_group AS
(
    SELECT subject_id, hadm_id,
    CASE
```

```sql
...
Reproducibility

- We recommend sharing your code
- Include a readme explaining how your code can be used.
- Include a license, so people know their rights to reuse.
Widely used internationally

Research

Education
A “datathon” model to support cross-disciplinary collaboration

Jerôme Aboab¹,*  Leo Anthony Celi¹  Peter Charlton¹  Mengling Feng¹  Mohammad Ghassemi¹  Dominic C. Marshall¹  Louis Mayaud¹  Tristan Naumann¹  Ned McCague¹  Kenneth E. Paik¹  Tom J. Pollard¹  Matthieu Resche-Rigon¹  Justin D. Salciccioli¹  and David J. Stone²,³

Author Affiliations

* Corresponding author. E-mail: dominic.marshall@ucl.ac.uk

All authors contributed equally to this work.

Science Translational Medicine  06 Apr 2019
Vol. 8, Issue 333, pp. 333ps8
DOI: 10.1126/scitranslmed.aad9072
Collaborative research

MIMIC is an openly available dataset developed by the MIT Lab for Computational Physiology, comprising deidentified health data associated with ~40,000 critical care patients. It includes demographics, vital signs, laboratory tests, medications, and more.

http://mimic.physionet.org