MIMIC-III
A Freely Available Critical Care Database

Tom J. Pollard

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
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
MIMIC-I

- 1992 - 1999
- Required consent from patient and health care provider
- 90 records, 40 hour duration
- Manual extraction of paper records for clinical data
MIMIC-III

- 2001 - 2012
- Waived consent for data collection
- ~40,000 patients
- Data extracted from digital systems
Accessing MIMIC

Two key steps to gaining access to MIMIC:

• **complete a recognized 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
Events tables

- **chartevents**: Charted observations for a patient
- **labevents**: Lab measurements both within hospital and outpatient clinics
- **inputevents**: Input fluids (e.g. intravenous medications)
- **microbiology events**: Microbiology measurements and sensitivities
- **noteevents**: Deidentified patient notes
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
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...
<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></td>
</tr>
<tr>
<td>Creatinine, mg/dL</td>
<td>0.7 0.7</td>
<td></td>
</tr>
<tr>
<td>White blood cell, K/uL</td>
<td>9.1 12.4</td>
<td></td>
</tr>
<tr>
<td>Neutrophil, %</td>
<td>37 37</td>
<td></td>
</tr>
<tr>
<td>GCS: Verbal</td>
<td>Confused</td>
<td></td>
</tr>
<tr>
<td>GCS: Moto</td>
<td>Obeys commands</td>
<td>Confused</td>
</tr>
<tr>
<td>GCS: Eye</td>
<td>Spontaneously</td>
<td>To speech</td>
</tr>
<tr>
<td>GCS: Verbal</td>
<td>Confused</td>
<td></td>
</tr>
<tr>
<td>GCS: Moto</td>
<td>Obeys commands</td>
<td>Confused</td>
</tr>
<tr>
<td>GCS: Eye</td>
<td>Spontaneously</td>
<td>To speech</td>
</tr>
<tr>
<td>Platelet, K/uL</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Creatinine, mg/dL</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>White blood cell, K/uL</td>
<td>16.8</td>
<td></td>
</tr>
<tr>
<td>Neutrophil, %</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>

**Laboratory Values**

- **Platelet, K/uL**: 48, 53
- **Creatinine, mg/dL**: 0.7, 0.7
- **White blood cell, K/uL**: 9.1, 12.4
- **Neutrophil, %**: 37, 37
- **GCS: Verbal**: Confused, Obeys commands, Spontaneously
- **GCS: Moto**: Confused, Obeys commands, Spontaneously
- **GCS: Eye**: Confused, Obeys commands, Spontaneously

**Medications**

- **Morphine Sulfate**
- **Vancomycin (1 dose)**
- **Piperacillin (1 dose)**
- **NaCl 0.9%**
- **Amiodarone**
- **Dextrose 5%**

**Doses**

- **10.0mL/hour**
- **1mg/min**
- **0.5mg/min**
- **50mL/hour**
- **25mL/hour**

*Graphs showing heart rate, O2 saturation, NIBP, mean, respiratory rate, intake volume, and output volume.*
Widely used internationally

**Research**

Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study

*The Lancet Respiratory Medicine*

Dr Romain Pirracchio, MD, Maya L Petersen, MD, Marco Carone, PhD, Matheiu Resche Rigon, MD, Prof Sylvie Chevret, MD, Prof Mark J van der Laan, PhD

Published Online: 23 November 2014

**Education**

BIOMEDIN 215 DATA DRIVEN MEDICINE

With the spread of electronic health records, increasingly large data repositories of clinical and other patient derived data are being built. These databases are large and difficult for any one specialist to analyze. To find the hidden associations within such data, we review methods for large-scale data-mining on electronic medical records, methods in natural language processing and text-mining of medical records, methods for using ontologies for notes.
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

†Corresponding author. E-mail: dominic.marshall@ch.embnet.org
* All authors contributed equally to this paper.

Science Translational Medicine 06 Apr 2015
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