Sharing is Caring: Moving Toward a Global Database for Critical Care

Ryan D. Kindle, MD
Research Affiliate, MIT Laboratory for Computational Physiology
Clinical Research Fellow, MGH/BIDMC/Harvard PCCM Program

January 20, 2020
Digitalization & Infectious Diseases 2020
Basel, Switzerland
Funding Disclosures

• National Institutes of Health  
  • R01-EB017205 and R01-EB001659  
• Philips Healthcare  
• Google  
• IBM
PhysioNet

The Research Resource for Complex Physiologic Signals

Data  Software  Challenges  Tutorials
BIDMC: MIMIC-III

- 61,532 ICU admissions
- 46,520 patients (7,870 neonates)
- Admitted 2001-2012 (year shifted)
- High-resolution data including:
  - Vitals signs, Labs, Imaging
  - Patient notes, Billing codes
- Post-hospitalization mortality

BIDMC: MIMIC-III MIMIC-IV

• Nearly 100,000 ICU admissions
• 2008-2018 (with admission year identified)
• More non-patient identifiers, e.g. specimen ID
• More timing information
  • Better support for real-time applications
• Richer pharmacy data (from eMAR)
  • Dose, route, frequency, duration, admin time
BIDMC: MIMIC-III MIMIC-IV

MIMIC-ED:
- ~200k patients

MIMIC-CXR:
- >377k images
- >64k patients
- Structured labels
- Free-text report
Philips Healthcare: eICU-CRD

- 208 hospitals across the US
  - 19 (9%) teaching hospitals
  - 335 intensive care units

- ~200,000 ICU admissions
  - ~140k unique patients

- Discharged in 2014 or 2015
Crowdsourcing

MIMIC-OMOP

This repository contains an Extract-Transform-Load (ETL) process for mapping the MIMIC-III database to the OMOP Common Data Model. This process involves both transforming the structure of the database (i.e., the relational schema), but also standardizing the many concepts in the MIMIC-III database to a standard vocabulary (primarily the Athena Vocabulary, which you can explore here).

DOCUMENTATION

- Resources
  - Achilles
  - OMOP Data Model
  - MIMIC Data Model

"WHERE IS ..."

Below in the README, we provide two sections. The first section, OMOP TABLES LOADED, lists the OMOP tables which have been populated from MIMIC-III. You can use this section to figure out what data generated each OMOP TABLE. For example, we can see that the OMOP CDM table person was populated using data from the patients and admissions table in MIMIC-III.

The second section, MIMIC TABLES EQUIVALENCE, lists all the tables in MIMIC-III, and shows where the data now exists in the OMOP CDM. For example, we can see that the MIMIC-III table patients was used to populate the OMOP CDM tables person and death.
Outreach: Datathons

• >30 events on 6 continents

• 36-48 hour collaborations of clinicians & data scientists

• Create interest & expertise to develop local databases
  • China, Korea, Singapore …
## Outreach: Datathons

<table>
<thead>
<tr>
<th>Starting Date</th>
<th>Event / Sponsors</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/31/2020</td>
<td>Health Datathon, Instituto Clinico Humanitas, Politecnico de Milano, European Society of Intensive Care Medicine</td>
<td>Milan, Italy</td>
</tr>
<tr>
<td>2/14/2020</td>
<td>Health Datathon, Society of Critical Care Medicine</td>
<td>Orlando, Florida, USA</td>
</tr>
<tr>
<td>3/27/2020</td>
<td>Health Datathon, Tokyo Medical &amp; Dental University, JSICM</td>
<td>Tokyo, Japan</td>
</tr>
<tr>
<td>April 2020</td>
<td>Health Data Science Symposium</td>
<td>Lhassa, Tibet</td>
</tr>
<tr>
<td>May 2020</td>
<td>Health Datathon, University of Puerto Rico</td>
<td>San Juan, Puerto Rico</td>
</tr>
<tr>
<td>June 2020</td>
<td>Health Datathon, Brazil and Argentina Critical Care Societies</td>
<td>Rio de Janeiro, Brazil</td>
</tr>
<tr>
<td>July 2020</td>
<td>Health Datathon, National University of Singapore</td>
<td>Singapore</td>
</tr>
<tr>
<td>Sept. 2020</td>
<td>Health Datathon, KoNECT, Seoul National University, Ajou University</td>
<td>Seoul, Korea</td>
</tr>
<tr>
<td>Nov. 2020</td>
<td>Health Datathon, Cleveland Clinic Abu Dhabi &amp; NYU Abu Dhabi</td>
<td>Abu Dhabi, UAE</td>
</tr>
</tbody>
</table>
Outreach: Health Ministries (Chile, Bhutan, NZ)

AI, big data could be key to improving Māori health

Artificial intelligence could be the key to lifting health outcomes for Māori. The Hack Aotearoa conference, held in Auckland, explores how predictive data, robotics, and new smart technologies can help to develop better health and wellbeing outcomes for New Zealanders – particularly Māori.

According to Dr Marise Stuart and Dr Materioria Lyndon, Māori feature strongly in negative health statistics in New Zealand.

They say that big data could be harnessed to ‘hack’ the health system, empower wellness, and lead to improved health outcomes across Aotearoa, with a priority on improving Māori health.

Leading researchers including Professor Eric Topol (Scripps Institute) and Dr. Leo Celli (MIT/Harvard) in the Artificial Intelligence (AI) sector, are coming to New Zealand to work alongside leading New Zealand data scientists and doctors.

They will also give presentations on the latest innovations and insights in health data science, as well as bring a key focus on how personalised medicine will be achieved, especially through the development of advanced digital tools.

Their aim is to explore the potential for Aotearoa to be a world indigenous leader in the fields of health and medicine by integrating Māori Tikanga with AI Technology.

The ‘Quadruple Aim of Healthcare’: To improve the health of populations, improve patient and workforce experience, and reduce healthcare costs also shape how new technologies could be used.
Multinational Federated Critical Care Database
Federated Database: Process

• Identify partner organizations & obtain institutional buy-in

• Create a de-identified local database analogous to MIMIC

• Develop a Common Data Model for data harmonization

• Generate protocols for sharing data directly or indirectly
Federated Database: Issues

• Multi-vendor systems: BIDMC != Epic != Cerner != Equipment

• Electronic Health Record implementations vary

• De-identification differs by language

• Privacy & data transfer regulations vary by country/institution
Federated Database: Brazil Development

• Local equivalent of MIMIC developed by a multi-disciplinary team of clinicians, researchers, data scientists, and managers

• Planned & executed each step of database construction with iterative evaluation of data fidelity as elements incorporated

• Work occurred in ICU space to ensure ease of collaboration
Federated Database: Brazil De-identification

• Unstructured data elements: Admission summaries, nursing notes, pharmacy notes, discharge summaries, etc.

• Note annotation: 1 physician, 2 medical students, 1 nursing student, & 1 data scientist manually annotated 100 notes

• De-identification algorithm trained using recursive neural networks & named entity recognition to automate process
Federated Database: Brazil De-identification
Federated Database: Data Sharing

• Direct Data Transfer Permissible: Distribution to researchers via PhysioNet & Google Cloud Platform to allow consolidation of databases and direct analysis with unified queries
  • Planned compatibility with multiple SQL and NoSQL options

• Alternatives:
  • Hospitals individually review researcher-proposed queries and send summative results without the individual patient data
  • Federated learning models (e.g. gradient descent) without direct knowledge of the underlying data elements via BigQuery
Federated Database: Indirect Sharing Issues

Kindle R. Unpublished data.
Federated Database: Indirect Sharing Issues

• Possible Solutions
  • Shared concept development, institution-level implementation
  • Reversion to the mean, law of big numbers, “good enough”
  • Break down legal and cultural barriers to direct data sharing
  • Accept cost of decreased insight for greater privacy protection
Global Collaboration Example: GOSSIS

• GOSSIS Consortium established in 2016 to develop a family of open source severity of illness scores for critical care patients

Conclusions

• De-identified high-resolution data can be safely shared with the research community without impacting patient privacy

• Too much clinical data remains siloed limiting opportunities for data curation and chances for improving patient care

• Legal and cultural barriers to direct data sharing may require indirect data sharing methods, for which research is needed to ensure data fidelity is preserved during model evaluation
Acknowledgements

• Leo Anthony Celi
• Roger G. Mark
• Tom Pollard (MIMIC, eICU-CRD)
• Alistair Johnson (MIMIC, eICU-CRD)
• Lucas Bulgarelli (Federated Database)
• Jesse Raffa (GOSSIS)
• E-mail: rkindle@mit.edu