Comparative Effectiveness using MIMIC II clinical data

Una-May O’Reilly
Leader, AnyScale Learning for All Group
Computer Science and Artificial Intelligence Lab
MIT
ALFA: AnyScale Learning for All

Data Science

ALFA Group Introduction
BeatDB Project
Gigabeats Overview
MIMIC II Databases

Quick Links
- Getting Access
- Waveform Database Overview
- Clinical Database Overview
- Waveform Data
- Clinical Data
- MIMIC II Statistics
- MIMIC II Clinical Database FAQs

The MIMIC II (Multiparameter Intelligent Monitoring in Intensive Care) Databases contain physiologic signals and vital signs time series captured from patient monitors, and comprehensive clinical data obtained from hospital medical information systems, for tens of thousands of Intensive Care Unit (ICU) patients. Data were collected between 2001 and 2008 from a variety of ICUs (medical, surgical, coronary care, and neonatal) in a single tertiary teaching hospital. The MIMIC II Clinical Database contains clinical data from bedside workstations as well as hospital archives. The MIMIC II Waveform Database includes records of continuous high-resolution physiologic waveforms and minute-by-minute numeric time series (trends) of physiologic measurements. Many, but not all, of the Waveform Database records are matched to corresponding Clinical Database records (for more information, see Record Matching). The databases are thoroughly de-identified (all PHI has been removed and all dates have been changed).

Both databases are distributed freely via PhysioNet. There are no restrictions on access to the MIMIC II Waveform Database. Access to the MIMIC II Clinical Database is available to qualified researchers who obtain human subjects training and sign a simple data use agreement (see Getting Access).

RESOURCES
Do adult ICU patients with sepsis have better, worse or the same outcomes, measured in terms of length of ICU stay and 30-day mortality, when they are treated with or without diuretics?
Definition of Selection Bias

- In the regression context, selection bias occurs when one or more regressors is correlated with the residual term.
- Recall that the residual captures the effects of all omitted and imperfectly measured variables.
- Thus any regressors that are correlated with the unmeasured or mismeasured factors will end up proxying for them.

The Problem

- In observational studies, especially those based on secondary data, it is common for important factors to be left out, i.e., subsumed into the residual.
- If a regressor ends up proxying for those factors, we cannot interpret its estimated coefficient as the effect of that regressor per se, since it also captures part of the effect of the omitted or mismeasured variables.

Methods for Addressing Selection Bias in Observational Studies
Susan L. Ettner, Ph.D. Professor, Division of General Internal Medicine and Health Services Research, UCLA
Example of Selection Bias

We are considering the impact of treating or not treating with diuretics

- If patients who are sicker are treated with diuretics, then negative effects of the drug may be overstated
  - But, severity of illness is correlated with mortality and length of stay
Selection Bias is not central problem here

FUNNY HOW IT'S ONLY THE PEOPLE
I DISAGREE WITH WHO ARE BIASED

DON'T BE DISTRACTED!!
Problem of this Presentation

“Non-treatment decision day dilemma”*

Non-treatment subgroup a.k.a “D-Neg” or D-
Treatment subgroup “D-Positives” or D+

* a.k.a Non-treatment Covariate Reference Timepoint Dilemna
Assembling D+ covariates
Assembling D- Covariates

“Non-treatment decision day (covariate) dilemma”*
D- and D+ Decision Days

No treatment: D-

Treatment: D+

ICU day vs count
Problem of this Presentation

“Non-treatment decision day dilemma”*

Non-treatment subgroup a.k.a “D-Neg” or D-
Treatment subgroup “D-Positives” or D+

* a.k.a Non-treatment Covariate Reference Timepoint Dilemna
Presenting a Solution

1. How we conduct a single outcome analysis
   - Controlled for selection bias
   - Covariates assumed to be already assembled

2. How to generate a “statistical” cohort D*
   - Sample n D*-'s
   - Assemble D*- covariates
     • Use statistically similar decision days to D+
   - Assemble D+ covariates relative to 1st day of treatment
   - Record stratified outcomes from outcome analysis on D*

3. How we use repetitions of Step 2 and analyze all outcomes from 2.4

A demonstration with the diuretics study
Retrospective Clinical Effectiveness Outcome Analysis

- Propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates.
  - Rosenbaum and Ruben technique

- Quintiles have
  - matched covariates
  - Separation by increasing likelihood of getting diuretics

**Stratification**

<table>
<thead>
<tr>
<th>Quintile</th>
<th>D+</th>
<th>D-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Q2</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>Q3</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Q4</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Q5</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Stratified outcomes

<table>
<thead>
<tr>
<th>Quintile 3</th>
<th>D+</th>
<th>D-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of ICU Stay</td>
<td>X1</td>
<td>X2</td>
</tr>
<tr>
<td>30-day mortality</td>
<td>X3</td>
<td>X4</td>
</tr>
</tbody>
</table>

**Step 1**
Problem of this Presentation

“Non-treatment decision day dilemma”*

Non-treatment subgroup a.k.a “D-Neg” or D-
Treatment subgroup “D-Positives” or D+

* a.k.a Non-treatment Covariate Reference Timepoint Dilemma
Our Solution based on Statistical Sampling

Consider

\( \pi \): distribution of \( T \) in \( D^+ \)
\( n \): size of \( D^+ \) group

*Choose a statistically sampled cohort \( D^* \)

1. \( n \) \( D^- \)'s from all \( D^- \) group -> \( D^* \)
2. choose \( T \)'s for \( D^* \) to match distribution \( \pi \)
3. collect covariates of \( D^* \)'s using \( T \)'s
4. run propensity score stratification and collect outcome data for \( \{D^+ \cup D^*-\} \)
5. collect quintile outcome data

Return to * for \( k \) iterations

Compare all iterations’ quintile outcomes of \( D^+ \) vs \( D^* \)

Are they statistically different?
Insight into this Resolution

• Each iteration:
  – Propensity scoring creates quintiles:
    • Balanced on co-variates *AND*
    • Balanced on likelihood of administering Diuretics

• Because each iteration, the D- patients and the timepoints used to collect their covariates
  – Are a result of random sampling
iteration reduces the uncertainty of sampling
Description of Diuretics Study Cohort

- Cohort size: 2341 unique patients
- Unique ICU stays: 3503
- Number of total patient-days: ~33K
- Number of unique patients where diuretics was administered: 349
- After filtering for missing values, number of covariates used for propensity analysis: 22
  - Static: age, ethnicity, sex
  - Time varying:
    • 10 Elixhauser,
    • 3 days of Fluid Inputs, Fluid Outputs, Fluid Balance
  - ICU_stay day used as treatment decision day (T)
Outcome - Length of stay
50 iterations

D+

D−
Outcome Analysis - Length of stay

![Box plots showing the distribution of length of stay for different quarters and conditions.]

- Q1: D+ ✓, D- ✓
- Q2: D+ ×, D- ×
- Q3: D+ ✓, D- ✓
- Q4: D+ ✓, D- ✓
- Q5: D+ ✓, D- ✓
Outcome – Mortality
50 Iterations

D+

D-

Quintile

Quintile
Outcome Analysis - Mortality
Open questions we are addressing

Theoretical Foundations

• Sample size for non-diuretics
  - Should it inform the number of iterations

• Number of iterations
Wrap Up

• Modest example of the sort of statistical technique evidence-based medicine needs
  \- Goal is to reduce sample noise, improve reliability of conclusions
    \- toward methodology standardization across studies

• Meta-analyses are important,
• Findings on different datasets need to be juxtaposed
• Data sharing to get big data sets is important
• MIMIC’s role cannot be over-exaggerated!
  \- +LCP, Roger Mark and Leo Celi

• What if?
What if?

• We could reduce the 6-8 months (or longer?) for any research team to
  • posit a hypothesis,
  • assemble and analyze the data
  • draw a conclusion
  – for a realistically large and messy set of observational clinical data,
  – to take less than a day.

• This is achievable with
  – data being organized with right abstractions, (generalizable, loss-less yet concise)
  – in-memory data assembly and cleaning methods (really fast)
  – scalable computational methods

Example: Scaling our Clinical Effectiveness/Outcome Analysis
  – soliciting other datasets that
    • Share same information potential wrt to same kind of outcome analyses?
      – Treatment times are not synchronized, need covariate matching
    • E.g. Clinical outpatient visits?
Acknowledgements

• The Diuretics Pilot Study Investigation Team
  – Kalyan Veeramachaneni, PhD
  – Daniele Ramazotti, M.Eng
  – Brian Bell, SB
  – Chris Moses, SB
  – Leo Celi, M.D.
  – John Huntziger, M.D.

• Funding
  – The Li Ka Shing Foundation
Covariate Reference Timepoint Selection

- **RCT**: controls D+/D- groups by randomization AND synchronizes timing of treatment
  - Half of group receives placebo/control -> D-
  - Half of group receives drug/case -> D+
  - Data has a common/synchronized reference timepoint -> T
  - “day of diuretics decision”
  - “first day of treatment”
  - Covariates are collected *relative* to T
    - SAPs at day T,
    - SOFA at T-1,
    - fluids-input at T-2
    - etc
Covariate Reference Timepoint Selection

But...MIMIC data is observational

• D-’s never got a placebo

• So how do we synchronize D-’s with D+’s around a reference time point?
  – Could synchronize everyone at same day after ICU admission, but...
Reference Timepoint Dilemma

Day 1
Covariate Reference Timepoint Selection

But...MIMIC data is observational

- D-’s never got a placebo
- So how do we synchronize D-’s with D+’s around a reference time point?
  - Could synchronize everyone at same day of ICU admission
    - but then first day of diuretics is not synchronized
  - Could synchronize everyone at latest day of diuretics decision
Timepoint Dilemma

Day 1

-2  -1  4

-4  -3  -2  -1  6  7  8  9

2  3  4  5  6  7  8  9
Covariate Reference Timepoint Selection

But...MIMIC data is observational

- D-’s never got a placebo
- So how do we synchronize D-’s with D+’s around a reference time point?
  - Could synchronize everyone at same day of ICU admission
    - but then first day of diuretics is not synchronized
  - Could synchronize everyone at latest day of diuretics decision
    - But then we have to make a poor choice for D-’s

Neither option is ideal

How to proceed so that studies are quick, systematic, and well founded statistically?
Outcome – Mortality
250 Iterations

% Deaths

Quintile

D+

Please check