Mapping and Visualizing Demographic Information in Structured and Unstructured Clinical Data

Clair Kronk and Danny T. Y. Wu
Tuesday, November 13, 2018
• We have no conflicts of interest to disclose.
Overview

• Problem
• Design
• Literature-based Example
• Data-driven Example
• Summary and Future Directions
Playing Both Sides: Visualizing Structured and Unstructured Data

• Why do we care?
  • Two patients with same diagnosis might require vastly different treatment regimens or prognoses
  • Such determinants might be in clinical notes

• Why record structured and unstructured data?
  • Medicine is often unpredictable
  • Doctors are human

• Why do we visualize?
  • Easier to process
  • Can be processed more quickly
  • Summarizes large volumes of data
Social Determinants of Health (SDOH)

- Conditions and circumstances that affect health risks and outcomes
- Examples include:
  - Income level
  - Educational opportunities
  - Gender inequity
  - Racial segregation
  - Availability of transportation

**Figure 2**
Impact of Different Factors on Risk of Premature Death

Gender Identity as SDOH

• Gender Identity
  • “One's innermost concept of self as male, female, a blend of both or neither – how individuals perceive themselves and what they call themselves. One's gender identity can be the same or different from their sex assigned at birth.” (Human Rights Campaign)
  • ~1.4 million Americans identify as transgender
  • Cisgender = Not transgender

• Known health concerns include social stigma, abuse, harassment, neglect, physical/sexual violence, STI prevalence, substance misuse, depression, anxiety, suicidal ideation, etc.

• We considered breast cancer, suicidal ideation, and depression before deciding to focus on suicide and depression
A quick note...

• Note that this demo only includes transgender men and women and not non-binary or third gendered people (e.g. hijra, māhū) simply because the data sources utilized do not contain any such individuals.

• Recording other gendered people in electronic health records (EHRs) should be a priority moving forward as it has been shown that different risk factors are associated with non-binary people versus binary transgender people.
Constructing an Interface

- Text entry/dropdown for gender identity, sex assigned at birth, ICD-9 codes
- Age range slider
- Limit for quicker load times
- UMLS search uses terminology mappings from ICD-9 to other terminologies
- Can download as file in multiple formats
Literature-Based Example: Suicide

• Suicide rate among women was highest for ages 45-54 (10.3 per 100,000)
• Suicide rate among men with highest for ages 65+ (32.3 per 100,000)
• National Transgender Discrimination Survey found 46% of trans men had attempted suicide and 42% of trans women had attempted suicide.
• How does this affect prognosis and treatment?

Statistics and image derived from
https://www.sciencedirect.com/science/article/pii/S0924933817318357,
https://www.verywellmind.com/men-and-suicide-2328492,
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539867,
Real World Example: MIMIC-III

- Medical Information Mart for Intensive Care
- Patients admitted to critical care units at large tertiary care hospital
- Includes components of electronic health records (EHRs):
  - Diagnostic Codes (Structured)
  - Clinical Notes (Unstructured)
Case Study with MIMIC-III: Depression

- MIMIC-III includes 6 individuals with transgender-related ICD-9 codes
- 13 trans individuals are detectable using terminology derived from the Gender, Sex, and Sexual Orientation (GSSO) ontology.
- Most common diagnoses were mental health related (depression, ADHD) and substance abuse related (tobacco, cocaine, etc.)
- How does this affect prognosis and treatment?

<table>
<thead>
<tr>
<th>MIMIC-III</th>
<th>UC Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of Birth</td>
<td>No</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Yes</td>
</tr>
<tr>
<td>Keywords for Deriving Gender Identity</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Literature Versus MIMIC-III

• Mental Health America notes that women experience depression at roughly twice the rate of men
  • MIMIC-III data: 1.56 times depressive cisgender women (1,604 / 20,392) versus depressive cisgender men (1,319 / 26,115)

• Reisner et al. (2015) noted that transgender youth are four times more likely to experience depression than their cisgender peers
  • MIMIC-III data contained 3.67 times depressive transgender people (3 / 13) versus cisgender people (2,923 / 46,507)
User Feedback and Suggestions

• The challenge of leveraging population health and individual health care

• Additional customization options such as scaling and classification by multiple factors (gender and age, for instance)
  • Added seven transformations and two stratification methods following this comment
User Feedback and Suggestions

- Difficult to distinguish four groups due to imbalanced data
- Heatmap colors chosen for the visually impaired, but not considered intuitive
- Lack of direct relationship between the heatmap and the age distribution
- Age distribution doesn’t allow comparison between gender identities or assigned sexes
Mockup of Changes
### Trans (Men vs Women)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Count</th>
<th>Scale</th>
<th>Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>1200</td>
<td>1,700</td>
<td>64</td>
</tr>
<tr>
<td>Women</td>
<td>1700</td>
<td>1,200</td>
<td>59</td>
</tr>
<tr>
<td>Cis</td>
<td></td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Trans</td>
<td></td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

-100% -80% -60% -40% -20% 0% 20% 40% 60%

10 20 30 40 50 60 70
Future Directions

- **Data**: Discussing recruitment of trans* people and analyzing provided clinical notes
  - Larger sample size needed in order to calculate false positive rates

- **Ontology**: Refining the ontology used for identifying trans* individuals and terminology
  - Gender, Sex, and Sexual Orientation ontology, available for comment here: https://github.com/Superraptor/GSSO

- **System**: Moving away from R Shiny toward a more customizable system like Django or Ruby on Rails

- **Interface**: Prioritization of user feedback

- **Other**: Working on implementing links to other datasets (such as the CDC Wonder API)
  - Public health link
References


