Missing Data: An Analysis of Lacks and Leaks in Health Data Collection

By: Taussia Boadi and Nino Owens
Mentor: Dr. Julia Lynch, PhD
- Understand the political circumstances behind government data collection
- Understand how their data collection strategies around socio-demographics allow them to see or be blind to health inequities
Significance

Why do we care about health data collection?
Health inequities and disparities are seen through the collection and comparison of health status and mortality data
“Governments (decide to) lack certain categories of sociodemographic data when possessing (or reporting) these data would exacerbate a potentially explosive social cleavage.”

Lynch, 2022
Data Leaks

Phenomenon where governments fail to publicize certain routinely collected variables
How is Health Data Collected
US Health Data Collection

- Vital Records and Death Certificates
- Census
- Self-reported Surveys
- Medical Records
- Cohort Studies
- Disease Registries
Understanding and Analyzing Health Data
The process of extracting information from archival records, often data files from companies and organizations.
Source Requirements

- Data collected by or at the request of a unit of the central government
- Data collected serially at least 4 times
- Data concerns the entire population of the country, not subgroups
- Data collected concerns multiple health topics, not a single disease
Qualitative Coding

- The process of searching for and identifying relationships, connections, or trends in text, media, and other data items
### Table 2.
Joinpoint incidence trends (2001-2017) for the most common cancers, all ages, all racial/ethnic groups combined by sex and age group, for areas in the United States with high-quality incidence data

<table>
<thead>
<tr>
<th>Sex and cancersite or type</th>
<th>Trends in 2001-2017</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st segment</td>
</tr>
<tr>
<td></td>
<td>Years</td>
</tr>
<tr>
<td>All sites</td>
<td></td>
</tr>
<tr>
<td>Both sexes combined</td>
<td>2001-2004</td>
</tr>
<tr>
<td>Males</td>
<td>2001-2004</td>
</tr>
<tr>
<td>Females</td>
<td>2001-2003</td>
</tr>
<tr>
<td>Children (aged 0-14 y)</td>
<td>2001-2017</td>
</tr>
<tr>
<td>AYA (aged 15-39 y)</td>
<td>2001-2017</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>----</td>
<td>---</td>
</tr>
<tr>
<td>TB</td>
<td>US</td>
</tr>
<tr>
<td>TB</td>
<td>US</td>
</tr>
</tbody>
</table>
What We Look For

- Changes in variables coded
- Changes in language used to code variables
- Additions or removals of certain variables
- Where variables are coded
Occurrence of Ethnicity Variables in US Health Data Reports
Occurrence of Gender Variables in US Health Data Reports
Next Steps

- Continue coding the case studies for the United States, United Kingdom, Sweden, and France
- Research data linkage patterns
- Unearthing the motivations of various actors in the data collection process
Lessons Learned

- Expertise in qualitative coding
- Importance of communication and transparency
- Familiarity with health data collection
Acknowledgements

- Measuring Mortality Team
  - Dr. Julia Lynch, Diya Amlani, Elin Berlin, Gabriella Rabito, Ramsey Radwan, Michael Tu
- Joanne Levy
- LDI Staff
- SUMR Cohort
Questions?