The Ethics of Using Big Data in Medicine

Progress vs Protection

Christopher S. Lathan, M.D., M.S., M.P.H.
Chief Clinical Access and Equity Officer
Associate Director Dana-Farber Cancer Network
Deputy Associate Director of Engagement and Equity Dana-Farber/Harvard Cancer Center
Overview

Big Data and Inequity
  • Examples
  • Medical Hubris

How do we protect communities?
  • Prospective community engagement
  • Transdisciplinary Research
  • Call to Action
• Lung Cancer Screening underutilization continues

• Differential impact based on region, and race

• Impact of the Pandemic, worsening screening access and follow up

• Can we use technology to level the playing field?

• Targeting of historically marginalized communities continue (Black and Indigenous, blue-collar workers and LGBTQ communities)
Potential impact of Covid-19 on Future Cancer Outcomes

Figure S2. Potential Impact of the COVID-19 Pandemic on Future Cancer Outcomes

COVID-19 PANDEMIC
- Reduced access to care
  - Fear of infection
  - Reallocation of health care resources
  - Unemployment leading to financial insecurity & insurance loss
  - Shutdowns & social distancing

Prevention & Early Detection
- Delayed routine care
  - Preventative visits
  - Screening
  - Abnormal test follow-up
  - Symptom follow-up

Diagnosis
- Later-stage diagnosis
  - Lower probability of survival
  - Fewer treatment options
  - More intensive treatment

Treatment
- Delayed/Modified treatment
  - Postponed surgery, radiation, and chemotherapy
  - Less intense chemotherapy
  - Non-standard care

Coronavirus image courtesy of CDC.

©2021 American Cancer Society, Inc., Surveillance Research
Practical Applications of AI and Big Data

- Big Data has many applications in medical care
- Physicians tend to be overly optimistic about the use of technology
- Evidence based data and data mining can be complimentary or antagonistic. *What about patient reported data?*
- Historically excluded communities are vulnerable to exploitation
- There are some protections federally to protect communities (14th Amendment) but they are not strong

Fig. 3. Artificial intelligence (AI) in cancer medical imaging. Deep learning algorithms in healthcare begins with the gathering of large amounts of data. The curation of this data is then used in the screening of patients to make better data driven diagnosis. Patients can be screened with medical imaging and the presence of biomarkers for disease. Image analysis involves the identification of image of interest and the areas of the image that are important. The application of information from datasets as well as the results of patient screening results in automated detection of malignant tumours. Through classification of different tumours, the application of AI algorithm's will then allow for the use of specific treatments optimised for each individual patient [48].

Dlamini et. al. Artificial intelligence and Big data in Cancer and Precision Oncology, Comp Struct Biotech J 2020.
Bias is Ubiquitous

Structural racism and classism is not an individual act
It is hard to understand how our data is impacted by our structure. While AI algorithms are helpful, they have been shown to “inadvertently” discriminatory. Facial Recognition Vocal recognition

The use of AI algorithms and big data is often touted as objective, leading to governmental use. Often the bias is built in
Insurance, loans, IQ/personality test for jobs, criminal sentencing software

Historically Marginalized communities have a good reason to question the intention of the medical community regarding their data
- Lack of representation in cancer genome datasets
- Careless stewarding of data/tissue for marginalized communities
- Eurocentric focus in germline datasets (Dementia)
- "Race medicine"

Lung cancer screening is important & use of AI might be helpful but w/o real efforts to combat bias, and acknowledgement of the past, we will do the same damage

Intent is often separate from impact

True Transdisciplinary approaches are needed to deal with Inequity in Big Data
Prospective Opportunities

Ask these questions: *

Does the tool increase the agency and self-determination of historically marginalized groups?
Is the data representative and relevant to our actual society?
Do the participants have agency?
Was the community engaged from the start?

Consciously diverse transdisciplinary teams involved with the research
Cross cutting all areas: technical, clinical, translational, basic science, social science, and policy

*The only way to keep us from making the same mistake and getting lost again is to take a different path*
Transdisciplinary Research

Three Types of Cross-disciplinary Research

• Multi
  Scientists across disciplines working together, addressing separate research questions

• Inter
  • Attempt to transfer knowledge from one discipline to another
  • Health equity innovation has high levels of complexity that challenge this approach

• Trans
  Mode of collaboration investigators operate entirely outside their disciplines. “beyond and outside all disciplines” and forming its own intellectual space

How do we protect communities as we move forward with tool development for LDCT?

**Prospective community engagement**
- Have you engaged communities and advocates at the start of your data project?
- Engaged faith-based groups, large employers, community leaders and policy experts?
- Do you have the desire to correct algorithms and acknowledge the bias & unintended consequences?

**Build transdisciplinary Teams**
- Is your technical team inclusive (technical skills, perspective) and representative of marginalized communities?
- Do you want authentic transdisciplinary input into your Data sets and AI algorithms?
## Transdisciplinary Challenges

<table>
<thead>
<tr>
<th>Key Elements</th>
<th>Description of Key Elements</th>
<th>Important Strategies</th>
<th>Significant Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>respect towards multiple perspectives</td>
<td>institutional support for transdisciplinary approaches</td>
<td>labor and time-intensive</td>
</tr>
<tr>
<td>Boundary-spanner</td>
<td>PI to bridge different discipline boundaries</td>
<td>diverse team members</td>
<td>difficult to evaluate/find</td>
</tr>
<tr>
<td>Flexibility</td>
<td>multiple pathways of integration &amp; collaboration across discipline norms</td>
<td>cross-disciplinary training &amp; opportunities</td>
<td>fear research will not be perceived as rigorous enough</td>
</tr>
<tr>
<td>Trust</td>
<td></td>
<td>build it</td>
<td>more reasons not to</td>
</tr>
<tr>
<td>Communication</td>
<td>across various discipline-specific languages</td>
<td>shared language and goals in operationalizing the research</td>
<td>academic publishing organized around disciplines</td>
</tr>
<tr>
<td>Stability</td>
<td>stability across expertise and subjectivity</td>
<td>make and invest time to build collaborations</td>
<td>difficulty in assigning roles to team members/time</td>
</tr>
<tr>
<td>Complexity</td>
<td></td>
<td>understanding what advances and hinders collaborative research</td>
<td>the need to not define the problem of analysis too narrowly or broadly</td>
</tr>
</tbody>
</table>
• Medical research has often been used knowingly and unknowingly to categorize and subjugate, immigrants, poor and descendants of slaves in the US.

• Research has not been “agnostic” and has until very recently has overtly frowned upon social justice approaches in medicine.

• The legacy of this remains today Big data and AI are the new frontier. Have we remembered any lessons from the past?

• If you haven’t already, I recommend folks listen the 1619 podcast from NYT episode 4

“When the Bad Blood Started”