Machine Learning in Health Care: Too Important to Be a Toy Problem

Sherri Rose, Ph.D.
Associate Professor
Center for Health Policy
Center for Primary Care & Outcomes Research

Co-Director
Health Policy Data Science Lab

Stanford University

drsherrrirose.org
@sherrrirose

November 18, 2020
“Learning two fields takes, surprisingly, twice as long as learning one. But it’s worth the investment because you get to solve real problems for the first time.”

Barbara Engelhardt | Princeton

“In both private enterprise and the public sector, research must be reflective of the society we’re serving.”

Rediet Abebe | Harvard

“...behind every data point there is a human story, there is a family, and there is suffering.”

Nick Jewell | LSHTM & UC Berkeley
DATA
Electronic Databases

The increasing availability of electronic health information offers a resource to health researchers.
Electronic Databases

The increasing availability of electronic health information offers a resource to health researchers.

General usefulness of this type of data to answer targeted scientific research questions is an open question.
Electronic Databases

The increasing availability of electronic health information offers a resource to health researchers.

General usefulness of this type of data to answer targeted scientific research questions is an open question.
Electronic Databases

The increasing availability of electronic health information offers a resource to health researchers.

General usefulness of this type of data to answer targeted scientific research questions is an open question varies.
Electronic Databases

The increasing availability of electronic health information offers a resource to health researchers.

General usefulness of this type of data to answer targeted scientific research questions is an open question varies.

May need novel statistical methods that have desirable properties while remaining computationally feasible.
Health Care Claims Data May Be Useful For COVID-19 Research Despite Significant Limitations

Maimuna S. Majumder, Sherri Rose
GENERALIZABILITY
<table>
<thead>
<tr>
<th>Prediction</th>
<th>Clustering</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Generalizability</td>
<td></td>
</tr>
</tbody>
</table>

The diagram shows a table with three columns: Prediction, Clustering, and Inference. The row labeled Generalizability spans across the columns.
Generalizability

Prediction

Clustering

Inference

RANDOMIZED TRIAL

OBSERVATIONAL STUDY

TARGET POPULATION
Generalizability

Prediction

Clustering

Inference

RANDOMIZED TRIAL

OBSERVATIONAL STUDY

Irina Degtiar
PhD Student
Harvard

Austin Denteh, PhD
Assistant Professor
Tulane

TARGET POPULATION
<table>
<thead>
<tr>
<th>Prediction</th>
<th>Clustering</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Generalizability
Machine Learning for Prediction in Electronic Health Data
Sherri Rose, PhD
The machine learning researchers who develop novel algorithms for prediction and the clinical teams interested in implementing them are frequently and unfortunately 2 nonintersecting groups.
DATASET SHIFT
Chronic Conditions

- Congestive heart failure: 40.7%
- COPD + asthma: 52.3%
- Paraplegia: 57.4%
- Chronic kidney disease: 59.5%
- Seizure disorders: 69.4%
- HIV/AIDS: 85.2%
- Type I diabetes mellitus: 92.6%
FAMILIAR QUESTION, DIFFERENT PROBLEM
Plan Payment Risk Adjustment

- Over 50 million people in the United States currently enrolled in an insurance program that uses risk adjustment
- Redistribute funds based on health
- Encourage competition based on efficiency and quality
- Massive financial implications
Plan Payment Risk Adjustment

Over 50 million people in the United States currently enrolled in an insurance program that uses risk adjustment

- Redistribute funds based on health
- Encourage competition based on efficiency and quality
- Massive financial implications

\[ Y = \theta X \]
Plan Payment Risk Adjustment

Over 50 million people in the United States currently enrolled in an insurance program that uses risk adjustment

- Redistribute funds based on health
- Encourage competition based on efficiency and quality
- Massive financial implications

\[
Y = \theta X
\]
Plan Payment Risk Adjustment

Over 50 million people in the United States currently enrolled in an insurance program that uses risk adjustment:

- Redistribute funds based on health
- Encourage competition based on efficiency and quality
- Massive financial implications

\[ Y = \theta X \]

Spending outcome

Input vector
Plan Payment Risk Adjustment

Over 50 million people in the United States currently enrolled in an insurance program that uses risk adjustment

- Redistribute funds based on health
- Encourage competition based on efficiency and quality
- Massive financial implications

\[ Y = \theta X \]
Variable Selection and Upcoding

Reduced set of 10 variables 92% as efficient

A Machine Learning Framework for Plan Payment Risk Adjustment

Sherri Rose
Variable Selection and Upcoding

Reduced set of 10 variables 92% as efficient

“...results for the risk adjustment algorithms that considered a limited subset of variables...performed consistently worse across all benchmarks.”
FAIRNESS
WHODECIDESTHERESEARCHQUESTION?

WHOISINTHETARGETPOPULATION?

WHATDOTHEDATARE...

HOWWILLTHEALGORITHMBEASSESSED?
<table>
<thead>
<tr>
<th>Prediction</th>
<th>Clustering</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalizability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Who decides the research question?  
Who is in the target population?  
What do the data reflect?  
How will the algorithm be assessed?
Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

Joy Buolamwini

MIT Media Lab 75 Amherst St. Cambridge, MA 02139

Timnit Gebru

Microsoft Research 641 Avenue of the Americas, New York, NY 10011

JOYAB@MIT.EDU

TIMNIT.GEBRU@MICROSOFT.COM
Black Patients Miss Out On Promising Cancer Drugs

A ProPublica analysis found that black people and Native Americans are under-represented in clinical trials of new drugs, even when the treatment is aimed at a type of cancer that disproportionately affects them.

For the 31 drugs which populations are most at risk for the cancers treated?

- White:
  - For the 31 drugs how often was each population the largest group represented in clinical trials?
  - Black: None
  - Similar Risk: None
  - Other: None

Note: Drugs are labeled "Similar Risk" if black Americans are at least 80 percent as likely as white Americans to be diagnosed with the cancer treated. (Chen and Wong 2018)
Black Patients Miss Out On Promising Cancer Drugs

A ProPublica analysis found that black people and Native Americans are under-represented in clinical trials of new drugs, even when the treatment is aimed at a type of cancer that disproportionately affects them.

Research Letter

September 28, 2020

The Exclusion of Older Persons From Vaccine and Treatment Trials for Coronavirus Disease 2019—Missing the Target

Benjamin K. I. Helfand, MSc1,2; Margaret Webb, BA3; Sarah L. Gartaganis, MSW, MPH3; et al

Note: Drugs are labeled "Similar Risk" if black Americans are at least 80 percent as likely as white Americans to be diagnosed with the cancer treated.

Chen and Wong (2018)
Algorithmic Fairness

Typical algorithmic fairness problem in computer science has

- outcome $Y$
- vector $X$ that includes a protected class or sensitive attribute $A \subset X$

**Goal:**
Create estimator for $f(X) = Y$ while ensuring the function is fair for $A$

Common measures of fairness are based on the notion of group fairness, striving for similarity in predicted outcomes or errors for groups.
Algorithmic Fairness

Typical algorithmic fairness problem in computer science has

- outcome $Y$
- vector $X$ that includes a protected class or sensitive attribute $A \subset X$

**Goal:**
Create estimator for $f(X) = Y$ while ensuring the function is fair for $A$

Common measures of fairness are based on the notion of **group fairness**, striving for similarity in predicted outcomes or errors for groups
Changes in financing and organization of mental health care, not new treatment technologies, made the difference

“Improvements ... evolved through ... more money, greater consumer choice, and the increased competition among ... providers that these forces unleashed”
Mental Health and Substance Use Disorders (MHSUD)

Risk adjustment in the Marketplaces recognizes only 20% of enrollees with MHSUD

Individuals with MHSUD can be systematically discriminated against
# Large Gains in Group Fairness vs. OLS

<table>
<thead>
<tr>
<th>Regression Method</th>
<th>$R^2$</th>
<th>MHSUD Net Compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>12.4%</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>Net Compensation</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Weighted Average</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>Mean Residual Difference</td>
<td>12.8</td>
<td></td>
</tr>
<tr>
<td>Ordinary Least Squares</td>
<td>12.9</td>
<td></td>
</tr>
</tbody>
</table>

Fair regression for health care spending

Anna Zink, Sherri Rose
## Large Gains in Group Fairness vs. OLS

<table>
<thead>
<tr>
<th>Regression Method</th>
<th>$R^2$</th>
<th>MHSUD Net Compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>12.4%</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>Net Compensation</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Weighted Average</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>Mean Residual Difference</td>
<td>12.8</td>
<td></td>
</tr>
<tr>
<td>Ordinary Least Squares</td>
<td>12.9</td>
<td></td>
</tr>
</tbody>
</table>

Fair regression for health care spending

Anna Zink, Sherri Rose
Large Gains in Group Fairness vs. OLS

<table>
<thead>
<tr>
<th>Regression Method</th>
<th>$R^2$</th>
<th>MHSUD Net Compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>12.4%</td>
<td>-$46</td>
</tr>
<tr>
<td>Covariance</td>
<td>12.4</td>
<td>-46</td>
</tr>
<tr>
<td>Net Compensation</td>
<td>12.5</td>
<td>-232</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>12.6</td>
<td>-411</td>
</tr>
<tr>
<td>Mean Residual Difference</td>
<td>12.8</td>
<td>-1208</td>
</tr>
<tr>
<td>Ordinary Least Squares</td>
<td>12.9</td>
<td>-1872</td>
</tr>
</tbody>
</table>

Fair regression for health care spending

Anna Zink, Sherri Rose
Ethical Machine Learning in Health Care

Irene Y. Chen,\textsuperscript{1} Emma Pierson,\textsuperscript{2} Sherri Rose,\textsuperscript{3} Shalmali Joshi,\textsuperscript{4} Kadija Ferryman,\textsuperscript{5} and Marzyeh Ghassemi\textsuperscript{4,6}

\textsuperscript{1}Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA; email: iychen@mit.edu
\textsuperscript{2}Microsoft Research, Cambridge, MA, 02143, USA
\textsuperscript{3}Center for Health Policy and Center for Primary Care and Outcomes Research, Stanford University, Stanford, CA, 94305, USA
\textsuperscript{4}Vector Institute, Toronto, ON, Canada
\textsuperscript{5}Department of Technology, Culture, and Society, Tandon School of Engineering, New York University, Brooklyn, NY, 11201, USA
\textsuperscript{6}Department of Computer Science, University of Toronto, Toronto, ON, Canada

Problem Selection

Data Collection

Outcome Definition

Algorithm Development

Post-Deployment Considerations

Disparities in funding and problem selection priorities are an ethical violation of principles of justice.

Focus on convenience samples can exacerbate existing disparities in marginalized and underserved populations, violating do-no-harm principles.

Biased clinical knowledge, implicit power differentials, and social disparities of the healthcare system encode bias in outcomes that violate justice principles.

Default practices, like evaluating performance on large populations, violate beneficence and justice principles when algorithms do not work for sub-populations.

Targeted, spot-check audits and lack of model documentation ignore systematic shifts in populations risks patient safety, furthering risk to underserved groups.
POLICY AND PRACTICE
Can Your Hip Replacement Kill You?

By JEANNE LENZER  JAN. 13, 2018
Can Your Hip Replacement Kill You?

By JEANNE LENZER  JAN. 13, 2018

---

Why Medical Devices Aren't Safer

Austin Frakt  THE NEW HEALTH CARE  APRIL 18, 2016

Things sometimes go wrong with airbags, food and drugs, prompting recalls. It can also happen with medical devices, though you’d think lifesaving devices like heart defibrillators or artificial hips would be closely monitored.

But the data needed to systematically and rapidly identify dangerous medical devices are not routinely collected in the United States.
Can Your Hip Replacement Kill You?

By JEANNE LENZER JAN. 13, 2018

Your medical implant could kill you
By Jeanine Lenzer

December 16, 2017 | 12:08pm | Updated

Why Medical Devices Aren't Safer

Austin Frakt
THE NEW HEALTH CARE APRIL 18, 2016

Things sometimes go wrong with airbags, food and drugs, prompting recalls. It can also happen with medical devices, though you'd think lifesaving devices like heart defibrillators or artificial hips would be closely monitored.

But the data needed to systematically and rapidly identify dangerous medical devices are not routinely collected in the United States.
Can Your Hip Replacement Kill You?

By JEANNE LENZER  JAN. 13, 2018

Why Medical Devices Aren't Safer

By Austin Frakt
THE NEW HEALTH CARE  APRIL 18, 2016

Things sometimes go wrong with airbags, food and drugs, prompting recalls. It can also happen with medical devices, though you'd think lifesaving devices like heart defibrillators or artificial hips would be closely monitored.

But the data needed to systematically and rapidly identify dangerous medical devices are not routinely collected in the United States.
Cardiac Stent Policy Implications

Implications for patients, hospitals, manufacturers, and regulators.

- How can this information be incorporated into the patient’s decision-making process?
- Will hospitals reconsider their complex contracting with manufacturers to avoid poorer-performing devices?
- Should manufacturers consider pulling stents from the market?
- How should regulators respond to postmarket information that was not available at the time of device approval?
IN CLOSING
Examining racism in health services research: A disciplinary self-critique

Rachel R. Hardeman PhD, MPH
J'Mag Karbeah MPH

International Journal of Epidemiology

Intersections of machine learning and epidemiological methods for health services research

Sherri Rose
Does Your Algorithm Have a Social Impact Statement?

Responsibility
Explainability
Accuracy
Auditability
Fairness

fatml.org/resources/principles-for-accountable-algorithms
Acknowledgements

Funding:
NIH Director’s New Innovator Award (DP2-MD012722)
Laura and John Arnold Foundation
NIH R01-GM111339

More about Cite Black Women:
Founded by Dr. Christen A. Smith, citeblackwomencollective.org