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

drsherrirose.org
@sherrirose

November 18, 2020

H

ECONOMICS

A

POLICY

OUTCOMES

H

“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

EHR ≠ EHR

- BILLING CLAIMS
- CLINICAL RECORDS
  - IMAGING
- REGISTRIES
- SURVEYS
- GOVERNMENT SOURCES
- WEARABLE & IMPLANTABLE TECH
- DIGITAL
- NEWS MEDIA

EHR ≠ EHR

BILLING CLAIMS

IBM Watson Health

CMS

CLINICAL RECORDS

IMAGING

REGISTRIES

SEER

Mass-DAC Data Analysis Center

SURVEYS

United States Census Bureau

cahps

GOVERNMENT SOURCES

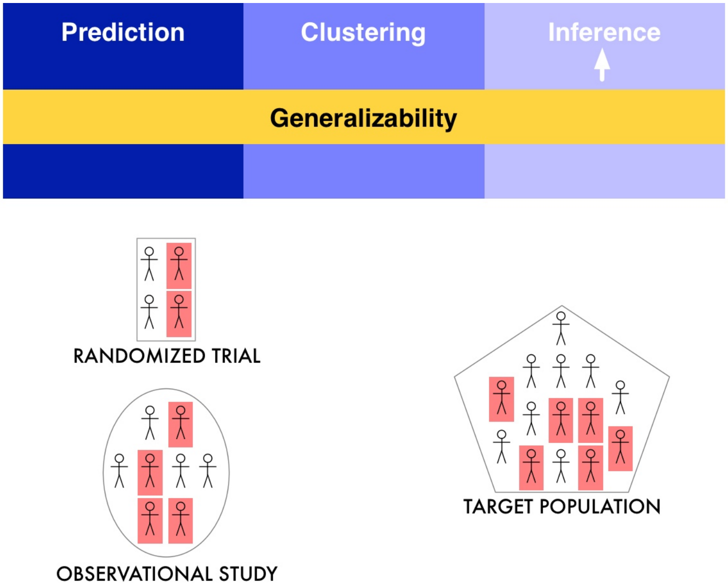
Health Affairs

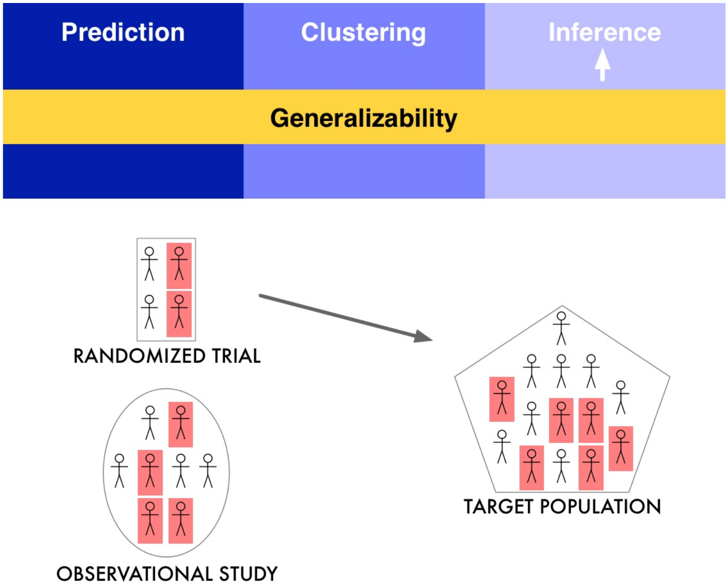
Health Care Claims Data May Be Useful For COVID-19 Research Despite Significant Limitations

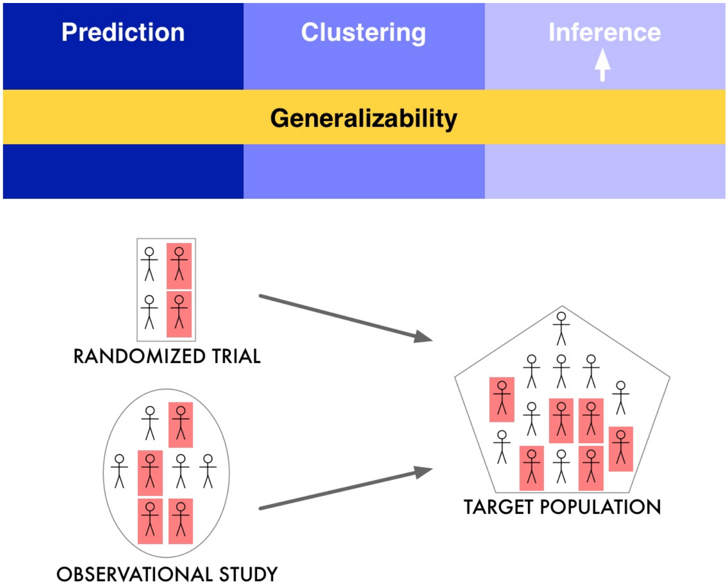
Maimuna S. Majumder, Sherri Rose

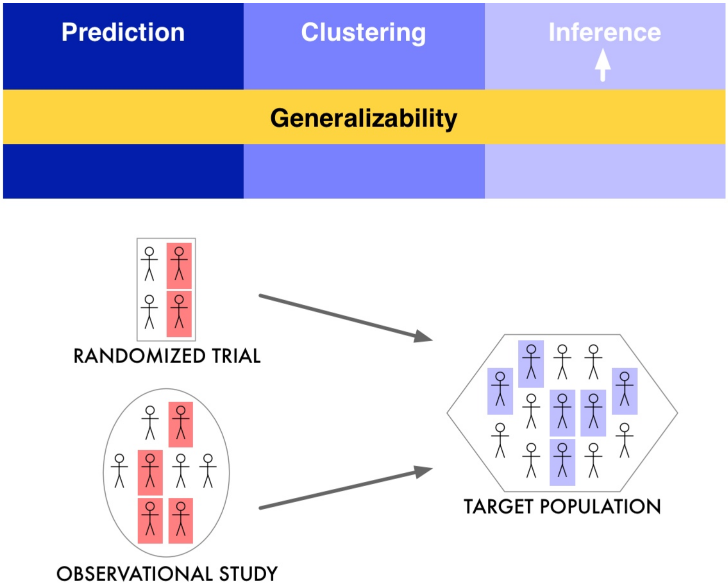
GENERALIZABILITY


Prediction
Clustering
Inference
Generalizability


Prediction
Clustering
Inference
Generalizability
RANDOMIZED TRIAL
OBSERVATIONAL STUDY
TARGET POPULATION

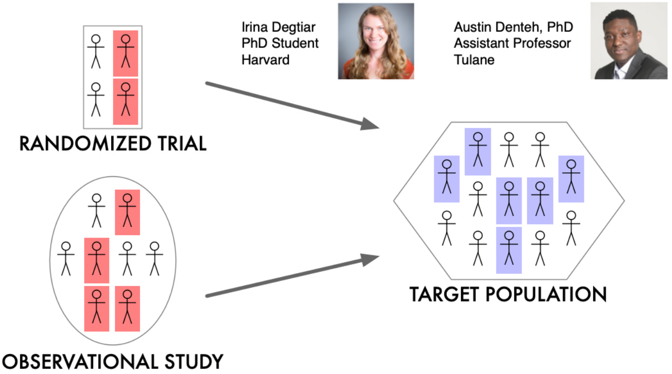

Prediction
Clustering
Inference
Generalizability
RANDOMIZED TRIAL
OBSERVATIONAL STUDY
TARGET POPULATION

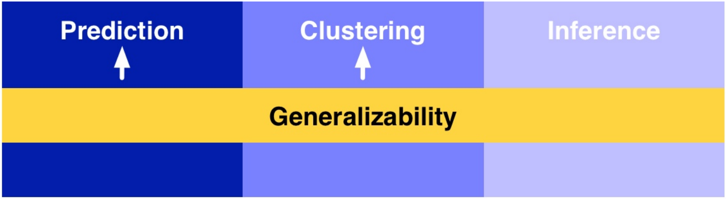

Prediction
Clustering
Inference
Generalizability
RANDOMIZED TRIAL
OBSERVATIONAL STUDY
TARGET POPULATION

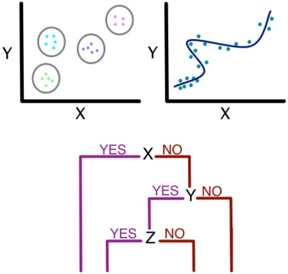

Prediction
Clustering
Inference
Generalizability
RANDOMIZED TRIAL
OBSERVATIONAL STUDY
TARGET POPULATION

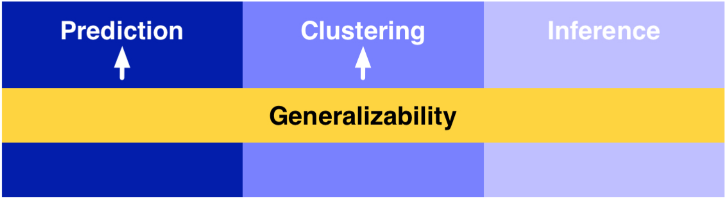
Prediction | **Clustering** | **Inference**

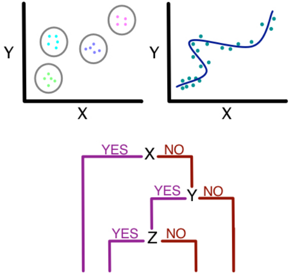
Generalizability




Prediction
Clustering
Inference
Generalizability


Y
X
Y
X
YES
X
NO
YES
Y
NO
YES
Z
NO



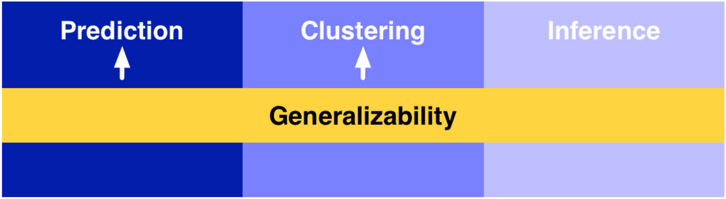


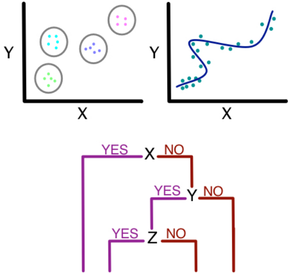
JAMA Network Open

Invited Commentary | Health Informatics

Machine Learning for Prediction in Electronic Health Data

Sherri Rose, PhD





JAMA Network | Open

Invited Commentary | Health Informatics

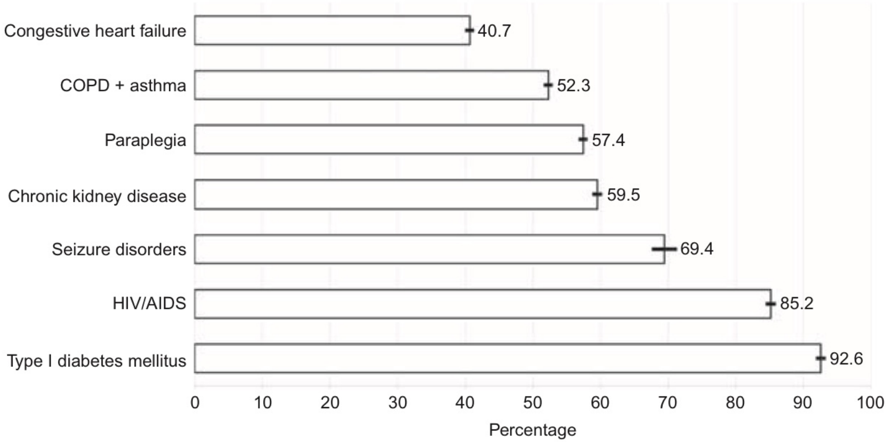
Machine Learning for Prediction in Electronic Health Data

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."

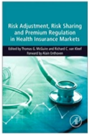
DATASET SHIFT

Chronic Conditions



Risk Adjustment for Health Plan Payment

Randall P. Ellis , Bruno Martins and Sherri Rose



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

Spending outcome

$$Y = \theta X$$

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

Spending outcome

$$Y = \theta X$$

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

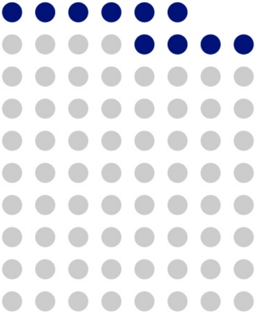
Spending outcome

$$Y = \theta X$$

Coefficient vector
Input vector

Variable Selection and Upcoding

Reduced set of 10 variables 92% as efficient

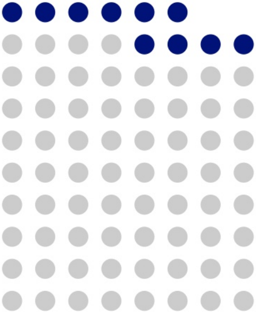


A Machine Learning Framework for Plan Payment Risk Adjustment

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."

Sample Selection for Medicare Risk Adjustment Due to Systematically Missing Data

Savannah L. Bergquist, Thomas G. McGuire, Timothy J. Layton, and Sherri Rose

A Machine Learning Framework for Plan Payment Risk Adjustment

Sherri Rose

FAIRNESS


Prediction
Clustering
Inference
Generalizability
Fairness

Who decides the research question?

Who is in the target population?

What do the data reflect?

How will the algorithm be assessed?

Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency

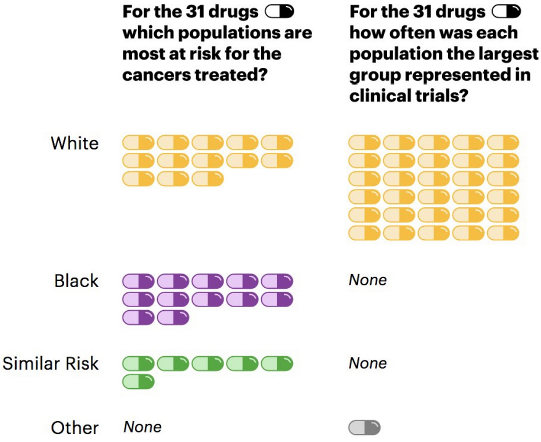
Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

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.



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.

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

For the 31 drugs how often was each population the largest group represented in clinical trials?

None

None

Research Letter

ONLINE FIRST FREE

September 28, 2020

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

Benjamin K. I. Helfand, MSc[1,2]; Margaret Webb, BA[3]; Sarah L. Gartaganis, MSW, MPH[3]; et al

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)

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

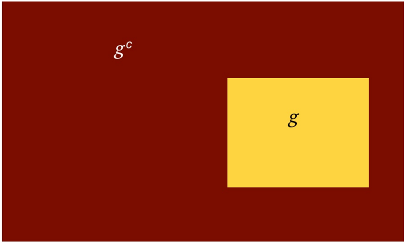
Typical algorithmic fairness problem in computer science has

- 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

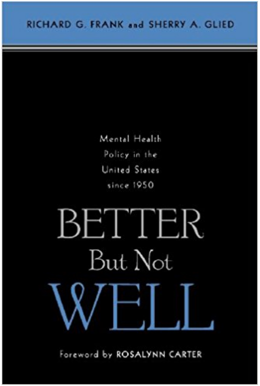
Common measures of fairness are based on the notion of **group fairness**, striving for similarity in predicted outcomes or errors for groups



Improving Mental Health Care, 1950-2000

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

By Ellen Montz, Tim Layton, Alisa B. Busch, Randall P. Ellis, Sherri Rose, and Thomas G. McGuire

Risk-Adjustment Simulation: Plans May Have Incentives To Distort Mental Health And Substance Use Coverage

Large Gains in Group Fairness vs. OLS

Regression Method	R^2	MHSUD Net Compensation
Average	12.4%	
Covariance	12.4	
Net Compensation	12.5	
Weighted Average	12.6	
Mean Residual Difference	12.8	
Ordinary Least Squares	12.9	

Biometrics JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

Fair regression for health care spending

Anna Zink, Sherri Rose

Large Gains in Group Fairness vs. OLS

Regression Method	R^2	MHSUD Net Compensation
Average	$\boxed{12.4\%}$	
Covariance	$\boxed{12.4}$	
Net Compensation	12.5	
Weighted Average	12.6	
Mean Residual Difference	12.8	
Ordinary Least Squares	12.9	

4%

Biometrics JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

Fair regression for health care spending

Anna Zink, Sherri Rose

Large Gains in Group Fairness vs. OLS

Regression Method	R^2	MHSUD Net Compensation
Average	12.4%	-$46
Covariance	12.4	-46
Net Compensation	12.5	-232
Weighted Average	12.6	-411
Mean Residual Difference	12.8	-1208
Ordinary Least Squares	12.9	-1872

98%

IBS Biometrics JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

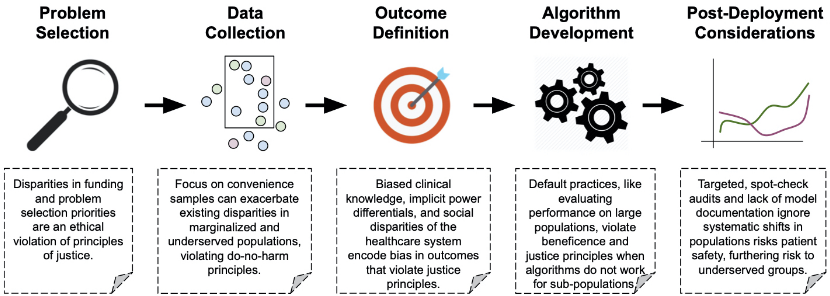
Fair regression for health care spending

Anna Zink, Sherri Rose

Ethical Machine Learning in Health Care

Irene Y. Chen,[1] Emma Pierson,[2] Sherri Rose,[3] Shalmali Joshi,[4] Kadija Ferryman,[5] and Marzyeh Ghassemi[4,6]

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



POLICY AND PRACTICE

Can Your Hip Replacement Kill You?

By JEANNE LENZER JAN. 13, 2018

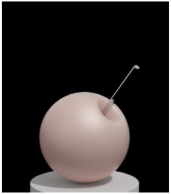
Can Your Hip Replacement Kill You?

By JEANNE LENZER JAN. 13, 2018

TheUpshot

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.

SundayReview | OPINION

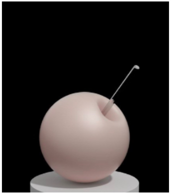
Can Your Hip Replacement Kill You?

By JEANNE LENZER JAN. 13, 2018

TheUpshot

Why Medical Devices Aren't Safer

Austin Frakt
THE NEW HEALTH CARE APRIL 18, 2016

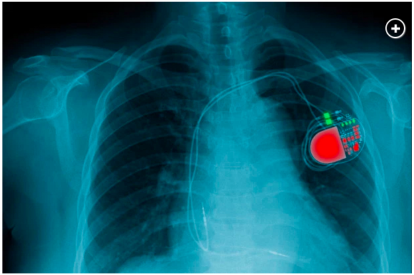


Things sometimes go wrong with airbags, food and drugs, prompting recalls. It can also happen with medical devices, though you'd think lifesaving devices like heart defibrillators or artificial hips would be closely monitored.

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

Your medical implant could kill you

By Jeanne Lenzer December 16, 2017 | 12:08pm | Updated



Shutterstock

SundayReview | OPINION

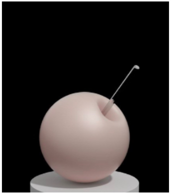
Can Your Hip Replacement Kill You?

By JEANNE LENZER JAN. 13, 2018

TheUpshot

Why Medical Devices Aren't Safer

Austin Frakt
THE NEW HEALTH CARE APRIL 18, 2016

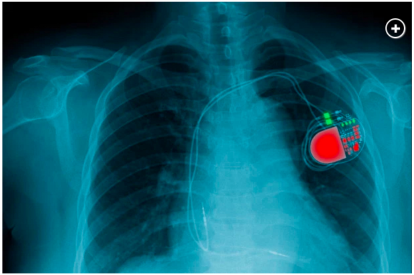


Things sometimes go wrong with airbags, food and drugs, prompting recalls. It can also happen with medical devices, though you'd think lifesaving devices like heart defibrillators or artificial hips would be closely monitored.

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

Your medical implant could kill you

By Jeanne Lenzer December 16, 2017 | 12:08pm | Updated



Shutterstock

npr set station news arts & life music programs shop

MEDICAL TREATMENTS

Are Implanted Medical Devices Creating A 'Danger Within Us'?

36:19

+ Queue

Download

Embed

Transcript

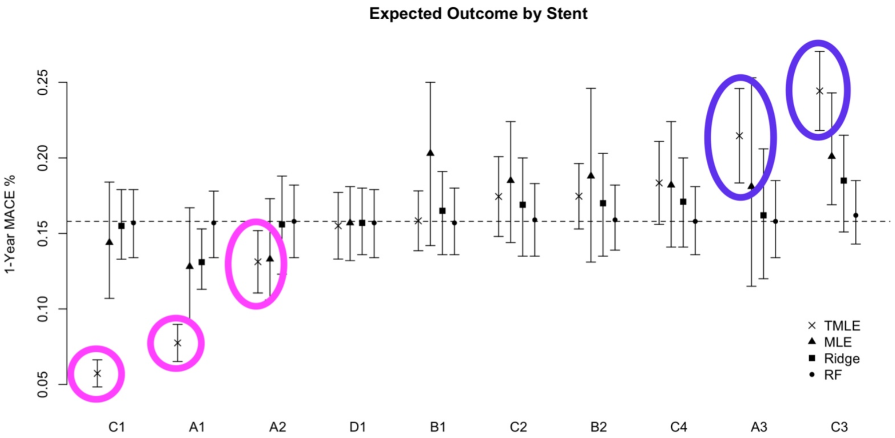
January 17, 2018 · 3:10 PM ET
Heard on Fresh Air

DAVE DAVIES

FRESH AIR

Medical journalist Jeanne Lenzer warns that implanted medical devices are approved with far less scrutiny and testing than drugs. As a result, she says, some have caused harm and even death.

Cardiac Stent Results



Biometrics JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

Double robust estimation for multiple unordered treatments and clustered observations: Evaluating drug-eluting coronary artery stents

Sherri Rose, Sharon-Lise Normand

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

HSR Health Services Research

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

Sam Adhikari, PhD
NYU

Austin Denteh, PhD
Tulane

Savannah Bergquist, PhD
Berkeley Haas

Akritee Shrestha, MS
Wayfair

Maia Majumder, PhD
Boston Children's/Harvard

Alex McDowell, PhD
MGH/Harvard

Toyya Pujol, PhD
Purdue

Anna Zink
Harvard

Irina Degtiar
Harvard

Noémie Sportiche
Harvard

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