## **Data Equity – An Introduction**

Jodi Lapidus, PhD

Presented as part of the **MICHR ABS Network Seminar Series** and the OCTRI-BERD Research Forum



## To frame today's discussion...

2

## Think about the last data project you were involved in or read about...

Click on the Jamboard link in the chat and share 🙂



Data equity as a journey...

### **Disclosures and Acknowledgements**

- I am presenting some of my own opinions, thoughts derived from my experience as:
  - a quantitative scientist

- a Professor in the OHSU-PSU School of Public Health
- a researcher and principal investigator who has engaged with native communities and organizations for over 25 years
- a person who is passionate about teaching others to effectively acquire and use data

- I will utilize information from resources published on the web and elsewhere, and I have made honest attempts to appropriately acknowledge
- I have no financial or other conflicts of interest to disclose.

## **Objectives for today's session**

#### Define data equity

- List frameworks that help us apply an equity lens to data-oriented projects
- Describe one or more data equity issues to consider in each facet/phase of a data-oriented project
- Articulate data equity considerations that arise in health studies
- Suggest strategies to inject data equity into projects that you may be involved in.

## A word about Data Projects

- We will use the words "research" and "study" inclusively in this session...
- Data Project → Project involving the acquisition and analysis of data designed to address a specific question.
  - Formal research studies
  - Evaluation Projects

- Strategic planning
- Needs assessments
- Surveillance efforts

## What is data equity? (1)

Ethos, philosophy or attitude toward using data for health, well-being, and equity<sup>a</sup>

Processes to help avoid racism, sexism, classism and other biases in our data and analysis<sup>b</sup>

References:

a. <u>www.communitycommons.org/collections/An-Introduction-to-Data-Equity</u>

b. <u>weallcount.com</u>

## What is data equity? (2)

7

- Consideration, through an equity lens, of the ways in which data is collected, analyzed, interpreted, and distributed.
- Addresses unequal opportunities to access data
- Acknowledges harm from data's misuse
- Raises the issue of data sovereignty, and the democratization of data.
- Pushes us to consider and mitigate the ways that data can reinforce stereotypes and exacerbate problems like racial bias.

Reference: www.jliconsultinghawaii.com/data-equity-training 8

Systematic error introduced into a study due to methods of sampling, measurement, or testing.

NorthwesternUS

Conscious. Unconscious.

Intentional. Accidental.

Undercount of American Indians in Undercount of American Indians in Hospital Discharge Databases in the Possibly (usually) a combination.

## Data Equity Frameworks (1)

- Frameworks that address equity systematically in each step of a data project.
- Ensure:

- study is designed,
- data is collected, analyzed, interpreted, and
- shared with diverse stakeholders,
- minimizing bias and avoiding exclusion
- Injecting an equity lens into each step impacts the narrative or the "data story" that is presented at the end.

## Data Equity Frameworks (2)

#### Community Information Exchange (CIE®) Data Equity Framework

Designed to facilitate strategic visioning when designing publichealth focused data systems

#### Do No Harm Project, Urban Institute

"how researchers and analysts can approach their work through a lens of diversity, equity, and inclusion"

#### References:

- ciesandiego.org/data-equity
- www.urban.org/projects/do-no-harm-project

## Data Equity Frameworks (3) We All Count



weallcount.com

**Project** Motivation



If you are involved in a project:

<u>Why</u> are you (your team) collecting, analyzing, and/or publishing data on a particular topic?

If you are a consumer of someone else's findings:

Can you easily ascertain <u>why</u> a certain study/project was done based on a report or an article?

Define core motivation and any secondary motivation(s) as specifically as possible



## **Project** Motivation

- High level: "We want to prevent injuries among AI/AN children"
- More specific: "Several NW tribes would like to design, implement and test the effectiveness of interventions to reduce children's injuries from motor vehicle crashes."

Example: Improving child passenger restraint use in six NW tribes (<u>Native Children Always Ride</u> <u>Safe – Native CARS</u>).





#### 14

### **Project** Motivation

Assess whether motivation(s) are compatible with equity goals

- Whose lens is most important?
- Who will <u>collect</u>, own, and analyze the data?
- Who will be responsible for interpreting and disseminating findings?
- Any potential conflicts between core and secondary motivations?



## Funding and Other Resources

- Having money to fund data projects means you get to find answers to a particular set of questions.
  - This presents an inherent equity issue.
  - Whose questions are being considered?
  - What data will we and won't we have? (Bias again!)
- Are any stakeholders financially (or otherwise) invested in a certain answer?.
  - Potential to skew results



## Funding and Other Resources

#### It is not all about funding...

- Can also be about other kinds of resources or expertise
- Can also be about *influence* political, social, emotional.

## How does funding interact with data and influence?





### Project Design

## Project Design is the phase where the WHY becomes the HOW...

## Project Design



- As you are designing (or working with a team to design) a data project, ask yourself:
  - Who is asking the questions? (Perspective)
  - What are the questions? (Motivation/Research Questions)
  - Where/How are you looking for answers? (Study Design/Data Sources)

Different **perspectives** with same **motivation** can lead to different **research questions** that require different **study designs and/or data sources** to address.

#### 20

## Project Design

- Perspective: Academic Researcher
- Motivation: Is child passenger safety improving in NW tribes?
  - Research Question: Have fewer AI children been hospitalized from motor vehicle crashes in areas where tribes implemented interventions vs. areas that did not?
- Data Source: Statewide hospital discharge data

- Perspective: Tribal organization
- Motivation: Is child passenger safety improving in NW tribes?
  - Research Question: Are more AI children using age- and size-appropriate passenger restraints while traveling in vehicles at/near NW tribes than in previous years?
- Data Source: Observational surveys of vehicles driving with child passengers



#### 21

## Project Design

- Perspective: Tribal residents
- Motivation: Is child passenger safety improving in NW tribes?
  - Research Question: Can we access no- or low-cost child safety seats more easily than before?

#### Data Sources:

- Tribal programs that distribute seats
- Focus groups with tribal parents and caregivers

- Perspective: Tribal Education
- Motivation: Is child passenger safety improving in NW tribes?
  - Research Question: Do school-age children know whether they should use a child safety seat or a seat belt?
- Data Source: Surveys of schoolchildren before and after passenger education program



### Project Design

Define your research question FIRST, then select your methodology.

"I want to conduct a randomized controlled trial to test different child passenger safety education programs at tribes."

Vs.

"To assess the effectiveness of child passenger safety education programs, <u>all</u> participating tribes will implement interventions via a cluster-randomized delayed-intervention strategy."



## Project Design



- Who sample and population?
- What what is the outcome variable? how granular is the measurement?
- Where what population and demographics? where in the geographic area and where in the population?
  - Whose geography? (<u>native-land.ca</u>)
- When at one point in time? before and after something? over a long period time?
- Why what type of question are you answering exploratory, descriptive, relationship, causal?

#### 24

## Data Collection & Sourcing



- Acquisition of data for our projects may consist of new data collection, or sourcing previously collected data.
- Data collection for a health-related data project can include one or more of the following:
  - Surveys, questionnaires or other standardized instruments
  - Qualitative data, such as elicitation interviews, focus groups
  - Observations
  - Physical/clinical measurements

## Data Collection & Sourcing

#### Data collection can be:

- Highly relational and social
- Embedded with subjectivity

#### Think about

- Influence and power dynamics of data collection methods
- Preconceived notions and hypotheses

- Some things to consider:
  - Who is selecting measurements?
  - Who is constructing categories?



- Whose definitions are being used? Do they apply to the current study?
- How are newly collected variables going to be used?
- For measurements from new technologies – how accurate and/or precise are they? Are they trusted? Are they accessible to all?
- What important features in "big data" sources do we need to be aware of?



#### 26

## Data Collection & Sourcing

- "Objective" data collection can easily become subjective...
- What categories get offered as options on a survey?
  - Response equity / inclusiveness
  - Equity in sampling
- Categorizing or collapsing for analysis
  - "Othering"
  - Balancing minority/small group privacy concerns
  - Reliability of estimates obtained

#### Do:

- Carefully consider categories before collecting and analyzing
- Report results multiple ways
- Be transparent about
  any dilemmas

#### Don't:

- Use the word "other". Choose more accurate terms for catch-all categories.
- Dismiss findings in small sample sizes with phrases like "not statistically significant".

#### weallcount.com

## Data Collection & Sourcing



- Data sourcing for a health-related data project can include :
  - Registry data
  - Administrative data e.g. health insurance claims
  - Electronic health record data
  - Surveys previously collected for another purpose
  - Well-characterized cohorts that have been followed long-term
- Some things to think about:
  - The purpose of the original vs. the motivation of your project.
  - What population was the original project meant to generalize to? Is it the same/different for your project?
  - What biases might be present in the dataset? (inclusion/exclusion criteria)
  - What is the sample size? Does your project require the whole dataset or a portion?



## Data Collection & Sourcing

#### Big Data

- Various definitions of "big data" but think in terms of datasets measured in TB instead of GB.
- Data "firehose"
- Wearable device data, Facebook posts, Tweets, Environmental sampling, etc.
- Inherently large samples can have:
  - Illusion of increased statistical power
  - Less than optimal data quality
  - Can have poor representativeness and substantial bias





## Data Equity Framework





## Analysis

#### Denominators are so important!

- Recall... Perspective can influence a project's motivation/research question, project design, as well as data collection and/or resourcing strategies for a study.
- It can also influence the unit of analysis (i.e. denominator on which our calculations are based)



Q: What is the representation of American Indians at my 3 data collection sites?



Q: What is the representation of American Indians at my 3 data collection sites?



Answer #1: 100% of the sites include American Indian participants.

Unit of analysis = site. Denominator = 3



Q: What is the representation of American Indians at my 3 data collection sites?



Answer #2: Average the %'ages of AI across sites (50% + 30% + 10%)/3 = 30%

Unit of analysis = Site. Denominator = 3



Q: What is the representation of American Indians at my 3 data collection sites?



Answer #3: # of AI divided by total number of people (1 + 3 + 2)/32 = 18%

Unit of analysis = People. Denominator = 32



34

Q: What is the representation of American Indians at my 3 data collection sites?



Answer #4: AI participant perspective, (50% + 30% + 30% + 30% + 10% + 10%)/6 = 26%

Unit of analysis = AI people only. Denominator = 6



Q: What is the representation of American Indians at my 3 data collection sites?



Answer #5: Non-Al participant perspective, (50% + 7\*30% + 18\*10%)/26 = 17%

Unit of analysis = Non-AI people only. Denominator = 26





...deciding

what variables

to include is

subjective and

should involve

stakeholders

Analyses, such as statistical models, can help show you how much impact each variable has on an outcome, but cannot tell you what variables to include in the first place!

#### **Anatomy of a Statistical Model**

Dependent Variable

Child using appropriate restraint (yes/no)

= Independent Variable Tribal interventions implemented? (yes/no)

Other Variables

- Age of child
- Relationship to driver
- Tribal law? (yes/no)
- Year of observation



## Anatomy of a Statistical Model

Analysis



## Better Practices for Modeling (<u>A Modeler's Manifesto</u>)

- In quantitative courses, we learn theory and methods and to apply sophisticated tools that can make analyses seem "automatic"
- As data scientists, we need to approach our practice more critically and less automatically.
- 1. Give greater context to scientific modeling
- 2. Greater collaboration in the modeling process
  - Triangulation with other data sources
  - Input from interdisciplinary teams
  - Uncertainty as invitation to dialogue
- 3. Emphasizing justice, ethics and engagement
  - Face impacts and implications throughout the phases of a project

Reference: datasciencebydesign.org/blog/writing-a-modelers-manifesto-for-more-transparent-ethical-data-science

- Analysis phase generates results.
  - Results are represented as numbers.
- Interpretation phase provides meaning and context to the numbers, as well as a narrative to explain them.

- Results:
- 40% of tribal elders reported falling in the year prior to the survey.
- Interpretation:
- Too many tribal elders are experiencing falls each year.

Must be careful to distinguish between the two.



## Interpretation



- When we interpret results, we are crafting a narrative or telling a story...
  - "This is what the results mean from this perspective"
- When interpreting results, we must:
  - Be honest and explain how we arrived at our interpretation(s)
  - Acknowledge our preconceived notions (e.g. hypotheses)
  - Collaborate with project stakeholders on interpretation
  - Explore alternative interpretations
  - Be careful not to over-state or over-generalize the results
  - Acknowledge limitations and/or what is still unknown
  - Take care that interpretations are not incurring harm to populations studied

Revisiting the core motivation for your data project should help frame interpretation.

40

/ Data Storytelling ~

Must be done

responsibly!



## **Communication & Dissemination**

- Communication and dissemination strategies planned through an equity lens require many voices!
  - Engage project stakeholders early!
  - Embrace lay language
  - Openness to alternatives
  - Focus on assets instead of deficits when appropriate
- Create transparent summaries of findings, but do not ignore issues of privacy

Of course, think about

ContentMediumAccess

What works well for one audience may not work at all for another.





## **Communication & Dissemination**

Tribal Councils

Residents of tribal communities

Academic Researchers

#### Content

- Language, Length, Depth, Tone, Perspective, Clarity, Complexity, Relevance, Cultural Translation
- Includes data visualizations and graphic representations

#### Medium

 Digital vs. Print, Static vs. Dynamic, Interactive, Branding, One-time vs. Updated, Group vs. Individual, etc.

#### 

 Technology requirements, Paywall vs. free, Ownership, Permanence, Training required, Copyright, Feedback options



## **Revisit Jamboard**

## The last **data project** I was involved in or read about was...



Data equity is a journey...

## Summary

- Considerations for planning and implementing studies or "data projects"
- Introduced concepts of data equity relevant to studies in the health sciences
- Highlighted one Data Equity Framework (weallcount.com)
  - Project Motivation
  - Funding/Resources
  - Project Design
  - Data Collection and Sourcing
  - Analysis
  - Interpretation
  - Communication & Dissemination
- Please think about how to apply the principles of data equity during your work

## Thank you for this opportunity

Feel free to reach out: lapidusi@ohsu.edu

Heartfelt appreciation to the Northwest Portland Area Indian Health Board and the native communities I have had the honor to engage with over the years.

Thank you to the organizers of:

- MICHR ABS Network Seminar Series
- OCTRI-BERD Research Forum
- and, all the attendees for spending your time!



SCHOOL OF PUBLIC HEALTH