BIAS IN BIG DATA
ADVANCING THE CONVERSATION ON SEXUAL AND GENDER MINORITY HEALTH
FRIDAY, JULY 19, 2019
11:00AM - 6:00PM CST

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
WORKSHOP OUTLINE

11:00 - 11:15  Registration
11:15 - 12:15  Introduction and Orientation
12:15 - 12:30  Break, Lunch
12:30 - 1:30  Yeshimabeit Milner, Data for Black Lives
1:30 - 2:00  Break, Check-in
2:00 - 3:00  Shifting the conversation to Health Research in SGM populations
3:00 - 4:00  How to promote the inclusion of diverse voices in big data and data science?
4:00 - 4:30  Where do we go from here?
4:30 - 6:00  Reception
USING SOCIAL MEDIA TO KEEP US CONNECTED

• We are livestreaming from @CONNECT_ISGMH

• Join the conversation with #BiasInBigData

• Our website will serve as a resource:
  https://isgmh.northwestern.edu/bias-in-big-data/
ENGAGE WITH US AND EACH OTHER

those joining us in Evanston - CONNECT@northwestern.edu

those joining us remotely – questions and comments via Bluejeans
1. **Be mindful in your participation.** Be a conscious observer of your own power in this conversation. Do not abuse that power.

2. **Make space.** We want this conversation to include as many voices as possible. Take the space you need to participate, but make space for others as well.

3. **Respect lived experiences.** The lived experiences of people in this room matter. Discussing bias and discrimination can be triggering and we are committed to engaging in discussion that is respectful and affirming.
SETTING THE STAGE FOR THE DAY
MICHELLE BIRKETT

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
SGM HEALTH DISPARITIES

• 2 to 3 times more likely to attempt suicide.¹
• More likely to be homeless.²,³,⁴
• Less likely to get preventive services for cancer.⁵,⁶
• Gay men are at higher risk of HIV and other STDs, especially among communities of color.⁷
• Lesbians and bisexual females are more likely to be overweight or obese.⁸
• Transgender individuals have a high prevalence of HIV/STDs,⁹ victimization,¹⁰ mental health issues,¹¹ and suicide¹² and are less likely to have health insurance than heterosexual or LGB individuals.¹³
• Highest rates of tobacco,¹⁴ alcohol,¹⁵ and other drug use.

⁷ Centers for Disease Control and Prevention (CDC). HIV among Gay and Bisexual Men [Internet]. Atlanta: CDC; 2017 Feb
Health is a function of the system
There are enormous disparities in HIV

![Diagram of Lifetime Risk of HIV Diagnosis among MSM by Race/Ethnicity]

- **African American MSM**: 1 in 2
- **Hispanic MSM**: 1 in 4
- **White MSM**: 1 in 11

Source: Centers for Disease Control and Prevention
Figure 1.2 - 2017 Rate of HIV Infection Diagnoses in Chicago by Community Area

Cases per 100,000 Population
- No Cases/Small Numbers (suppressed)
- 8.2 - 20.5
- 20.6 - 40.3
- 40.4 - 55.7
- 55.8 - 88.7
- High Economic Hardship in 2015

Community Areas
Most Impacted (Red)
Figure 1.2 - 2017 Rate of HIV Infection Diagnoses in Chicago by Community Area

Cases per 100,000 Population:
- No Cases/Small Numbers (suppressed)
- 8.2 - 20.5
- 20.6 - 40.3
- 40.4 - 55.7
- 55.8 - 88.7
- High Economic Hardship in 2015

Community Areas:
- Most Impacted (Red)
BIG PROBLEMS CAN’T BE SOLVED BY A SINGLE DISCIPLINE OR SECTOR
1) build and support transdisciplinary research collaborations

2) develop and foster methods for the capture, integration, and analysis of rich social data (dynamic, multilevel, multiplex, relational/network, spatial, population-based surveillance, text, and social media) that are relevant to health

3) develop deeper theoretical understandings of stigma and the systemic production of health disparities within stigmatized populations, and apply this knowledge to shape new intervention strategies.
GOALS FOR TODAY
GOALS FOR TODAY

CREATE A SPACE THAT WILL BRIDGE SOME BOUNDARIES
CONNECTING CAMPUSES

>50 Joining in-person

~20 Joining from Evanston

100 Unique Remote Connections
CONNECTING CAMPUS

UIC
LOYOLA UNIVERSITY CHICAGO
S
Harvard
University of Maryland
Drexel University
Emory University
NYU
Johns Hopkins University
PennState
University of California, Berkeley
University of California, Los Angeles
University of California, Santa Barbara
CONNECTING ACROSS SECTORS AND DISCIPLINES
CONNECTING ACROSS SECTORS AND DISCIPLINES

- Community Based-Organizations
- Healthcare services workers and care providers
- Nonprofit researchers
- Activists and advocates
BRINGING LIVED EXPERIENCE TO THE TABLE

• OPENING THE CONVERSATION
• MAKING VOICES HEARD
• IMPROVING DATA SCIENCE, IMPROVING LIVES
GOALS FOR TODAY

DISPEL THE MYTH THAT DATA IS UNBIASED
GOALS FOR TODAY

EDUCATE ON HOW DATA SYSTEMS AND AUTOMATED
DECISION MAKING AMPLIFY BIAS
SHOW THROUGH NUMEROUS EXAMPLES HOW ALREADY VULNERABLE PEOPLE ARE HURT BY DATA AND DATA SYSTEMS
WE HOPE TO HIGHLIGHT AND AMPLIFY THE GOOD WORK ALREADY BEING DONE
WE HOPE TO HIGHLIGHT AND AMPLIFY THE GOOD WORK ALREADY BEING DONE

GOALS FOR TODAY

FAT/ML

DATA FOR BLACK LIVES

AI NOW INSTITUTE

Data & Society
GOALS FOR TODAY

BRING THE CONVERSATION BACK TO SEXUAL AND GENDER MINORITY POPULATIONS
GOALS FOR TODAY

HOW DIVERSE VOICES CAN BE BROUGHT INTO DATA SCIENCE – BOTH BROADLY AND LOCALLY
GOALS FOR TODAY

A CONVERSATION – NOT A LECTURE
ACTIVITY TIME!
WHAT IS RESEARCH?
Identify Real World Problem
Identify Real World Problem
Create a Mathematical Model of Problem
Identify Real World Problem

Create a Mathematical Model of Problem

Receive Results
Identify Real World Problem
Create a Mathematical Model of Problem
Receive Results
Apply solution
1. Identify Real World Problem
2. Create a Mathematical Model of Problem
3. Receive Results
4. Apply Solution
Identify Real World Problem

Create a Mathematical Model of Problem

Receive Results

Apply solution
In 2008, researchers from Google explored this potential, claiming that they could “nowcast” the flu based on people’s searches. The essential idea, published in a paper in Nature, was that when people are sick with the flu, many search for flu-related information on Google, providing almost instant signals of overall flu prevalence. The paper demonstrated that search data, if properly tuned to the flu tracking information from the Centers for Disease Control and Prevention, could produce accurate estimates of flu prevalence two weeks earlier than the CDC’s data—turning the digital refuse of people’s searches into potentially life-saving insights.

[When explored] Google’s algorithm was quite vulnerable to overfitting to seasonal terms unrelated to the flu, like “high school basketball.”
Identify Real World Problem
Create a Mathematical Model of Problem
Receive Results
Apply solution
Identify Real World Problem

Create a Mathematical Model of Problem

Receive Results

Apply solution
Identify Real World Problem

Create a Mathematical Model of Problem

Receive Results

Apply solution
None of this is new – but in data science/big data – biases can be amplified at scale – and more easily hidden.
15 MINUTE BREAK – PICK UP A BOX LUNCH.
INTRODUCE YOURSELF TO TWO PEOPLE

Back at 12:30 CST with Yeshimabeit Milner, Data for Black Lives

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
YESHIMABEIT MILNER
Founder and Executive Director of Data for Black Lives

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
The color of risk.
risk

danger
jeopardy
peril
hazard
menace
threat
possibility
chance
gamble
probability

Thesaurus...
THE COMING OF THE SUPER-PREDATORS

By John J. Dilulio, Jr.
As Ex-Theorist on Young ‘Superpredators,’ Bush Aide Has Regrets

By ELIZABETH BECKER

PHILADELPHIA, Feb. 8 — From his perch as the director of the new White House Office of Faith-Based and Community Initiatives, which he believes will help uplift many needy people but particularly the most troubled teenagers, John J. Dilulio Jr. conceded today that he wished he had never become the 1990’s intellectual pillar for putting violent juveniles in prison and condemning them as “superpredators.”

“If I knew then what I know now, I would have shouted for prevention of crimes,” Mr. Dilulio said during an interview in the clubby University of Pennsylvania office that he is temporarily vacating to join the White House staff.

Instead, five years ago, Mr. Dilulio created a whole theory around the notion that “a new generation of street criminals is upon us — the youngest, biggest and baddest generation any society has ever known.”

“Based on all that we have witnessed, researched and heard from people who are close to the action,” he wrote with two co-authors, “here is what we believe: America is now home to thickening ranks of juvenile superpredators — radically impulsive, brutally remorseless youngsters, including even more preteenage boys, who murder, assault, rape, rob, burglarize, deal deadly drugs, join gun-toting gangs and create serious communal disorders.”

“At core,” the authors said, “the problem is that most inner-city chil-

 dropped by more than half.

“His prediction wasn’t just wrong, it was exactly the opposite,” said Franklin E. Zimring, professor of law at the University of California at Berkeley and director of the university’s Earl Warren Legal Institute.

“His theories on superpredators were utter madness.”

Mr. Dilulio still defends the quality of his research, saying that “the data we used was correct” — largely crime statistics and projections of growth in the teenage population. Of his conclusions, however, he said, “Thank God we were wrong.”

When pressed now on the subject of prisons, he argues for more federal money for church programs instead, and for ex-elonos as well as those programs to counsel children whose parents are behind bars. When he talks of offenders, he says that only “a certain fraction have to be incarcerated, which we do with a heavy heart.”

Back in 1996, he complained that “some prisons are virtual resorts.”

“There are, to be sure, good moral and cost-effectiveness arguments for scaling back prisoner amenities and services,” he wrote.

And as recently as last year a report by Human Rights Watch blamed the theory of superpredators for state initiatives to move juvenile offenders into the adult criminal justice system.

“I’m sorry for any unintended consequences,” Mr. Dilulio said today. “But I am not responsible for teen-

now. “It’s right here,” he said, slapping his 1997 Catholic catechism on the desk. “Prevention is the only reasonable way to approach these problems.”

Changeling or genuine convert? Professor Zimring lauds at trying to answer that question.

“There are areas where John Di-

luio has done great work,” he said. “He’s a very talented, enthusiastic person, and he has an important mission I fully support.”

Others are more critical. “The superpredator thing led to horrific legislation,” and “while he may have backed away from the idea, he has never really recanted it,” said Jerry Miller, president of the nonpartisan National Center on Institutions and Alternatives. “And that makes me nervous.”

But within the world of churches and other religious organizations trying to minister to the poor, there are many supporters of his, said Jim Wallis, founder of Call to Renewal, a national ecumenical group that engages in such ministry.

John moved from crime control to crime prevention when he went into the streets and fell in love with those kids,” Mr. Wallis said. “He encountered the poor, and he found his faith again in the face of our poor’s children.”

For his part, Mr. Dilulio said one advantage in his change of views was that it had brought attention to him that was now drawing a large audi-

John J. Dilulio Jr., the director of the new White House Office of Faith-Based and Community Initiatives, once warned of growing ranks of teenage “superpredators.” Then, he says, he had an epiphany.

Laura Pedrick for The New York Times

ology with his superpredator the-

ory. But although a respected aca-

demic, he was suddenly questioned by peers, who said he seemed to be providing cover for what they consid-

ered partisan politics.

“He became a sensationalist, a simplistic analyst who rather toad-

ied to that point of view,” said Norval Morris, professor of law at the Uni-

versity of Chicago and co-author of Penn, where he is a professor of government policy. “God had given me a Rolodex, good will and a pas-

sion that was sometimes misdirected,

and I knew that for the rest of my life I would work on prevention, on helping bring caring, responsible adults to wrap their arms around these kids.”

He tried, he said, to put the brakes a sheet of articles he published em-

phasizing churches over prisons, or opposing Congress’s welfare over-

haul as legislation that undercut the most vulnerable families.

He also took to the streets of Phila-

delphia to do firsthand research there and engage in community service teaching. And he promoted the ministries of Northeastern cler-

now.
Revisiting the ‘Crack Babies’ Epidemic That Was Not

This week’s Retro Report video on “crack babies” (infants born to addicted mothers) lays out how limited scientific studies in the 1980s led to predictions that a generation of children would be damaged for life. Those predictions turned out to be wrong. This supposed epidemic — one television reporter talks of a 500 percent increase in damaged babies — was kicked off by a study of just 23 infants that the lead researcher now says was blown out of proportion. And the shocking symptoms — like tremors and low birth weight — are not particular to cocaine-exposed babies, pediatric researchers say; they can be seen in many premature newborns.
The ‘Welfare Queen’ Is a Lie

Programs that should be crafted around people’s needs are instead designed to deal with a problem that doesn’t exist.

RACHEL BLACK, ALETA SPRAUGE, AND NEW AMERICA WEEKLY | SEP 28, 2016 | BUSINESS

Like The Atlantic? Subscribe to The Atlantic Daily, our free weekday email newsletter.

Email SIGN UP

At a campaign rally in 1976, Ronald Reagan introduced the welfare queen into the public conversation about poverty: “She used 80 names, 30 addresses, 15 telephone numbers to collect food stamps, Social Security, veterans’ benefits for four nonexistent deceased veteran husbands, as well as welfare. Her tax-free cash income alone has been running $150,000 a year.”
“When you have a combination of a society that protects racism, misogyny and transphobia, it creates insurmountable odds…”

Louis Mitchell
TransFaith
IN THE AGE OF BIG DATA,
IF WE ARE NOT AWARE OF THIS HISTORY,
WE RISK REPEATING IT.
HISTORY + VALUES

Objective Function
What are we optimizing?

Input → ALGORITHM → Output

DATA

RISK RATIO
CREDIT SCORE
CAR INSURANCE PAYMENT
MAP ROUTE
NETFLIX RECOMMENDATION
ZIP Code as a proxy for race

HOLC GRADING AND AREA INCOME
The Home Owner’s Loan Corporation (HOLC) identified neighborhoods by four categories/colors. The green areas were “Best”, while the red areas were “Hazardous”. Today, 74% of the “Hazardous” areas remain low-to-moderate income (LMI) nationally, while “Best” areas are 94% middle-to-upper income (MUI).
Here is the ranking for the city of Miami, FL:

<table>
<thead>
<tr>
<th>Grades/Income</th>
<th>LMI</th>
<th>MUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Best</td>
<td>7.36%</td>
<td>92.64%</td>
</tr>
<tr>
<td>B Desirable</td>
<td>26.09%</td>
<td>71.91%</td>
</tr>
<tr>
<td>C = “Definitely Declining”</td>
<td>34.16%</td>
<td>65.84%</td>
</tr>
<tr>
<td>D = “Hazardous”</td>
<td>34.16%</td>
<td>65.84%</td>
</tr>
</tbody>
</table>

MINORITY POPULATION AND GRADE
The examiners also graded neighborhoods by their racial/ethnic composition. The presence of minority communities resulted in downgrading under their system. 64% of the communities identified as “Hazardous” by the HOLC in the 1930’s have majority-minority populations in the 2010 Census.
Here is the ranking for the city of Miami, FL:

<table>
<thead>
<tr>
<th>Grades/Minority</th>
<th>White</th>
<th>Minority</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Best</td>
<td>43.87%</td>
<td>56.13%</td>
</tr>
<tr>
<td>B Desirable</td>
<td>21.36%</td>
<td>78.64%</td>
</tr>
</tbody>
</table>
Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk
Peers

Please think of your friends and the people you hung out with in the past few (3-6) months.

39. How many of your friends/acquaintances have ever been arrested?
   - None ☐ Few ☑ Half ☐ Most

40. How many of your friends/acquaintances served time in jail or prison?
   - None ☐ Few ☑ Half ☐ Most

41. How many of your friends/acquaintances are gang members?

54. How often do you have contact with your family (may be in person, phone, mail)?
   - No family ☐ Never ☐ Less than once/month ☐ Once per week ☑ Daily

55. How often have you moved in the last twelve months?
   - Never ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5+

56. Do you have a regular living situation (an address where you usually stay and can be reached)?
   - No ☐ Yes

57. How long have you been living at your current address?

73. What were your usual grades in high school?
   - A ☐ B ☑ C ☐ D ☐ E/F ☐ Did Not Attend

74. Were you ever suspended or expelled from school?
   - No ☐ Yes

75. Did you fail or repeat a grade level?
   - No ☐ Yes

76. How often did you have conflicts with teachers at school?

Thinking of your leisure time in the past few (3-6) months, how often did you have the following feelings?

95. How often did you feel bored?
   - Never ☐ Several times/mo ☐ Several times/wk ☐ Daily

96. How often did you feel you have nothing to do in your spare time?

The next statements are about your feelings and beliefs about various things. Again, there are no 'right or wrong' answers. Just indicate how much you agree or disagree with each statement.

127. "A hungry person has a right to steal."
   - ☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree

128. "When people get into trouble with the law it's because they have no chance to get a decent job."
   - ☑ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
<table>
<thead>
<tr>
<th>VAGINAL BIRTH AFTER CESAREAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height &amp; weight optional; enter them to automatically calculate BMI</td>
</tr>
</tbody>
</table>

Maternal age: 29 years

Height (range 54-80 in.): in

Weight (range 80-310 lb.): lb

Body mass index (BMI, range 15-75): 28 kg/m²

African-American?: no

Hispanic?: no

Any previous vaginal delivery?: no

Any vaginal delivery since last cesarean?: no

Indication for prior cesarean of arrest of dilation or descent?: no

Calculate

Predicted chance of vaginal birth after cesarean: **75.5%**

95% confidence interval: **[73.7%, 77.3%]**

This calculator is based on the equation published in the article "Development of a nomogram for prediction of vaginal birth after cesarean" cited below. It is designed for educational use and is based on a population of women who received care at the hospitals within the MFMU Network. Responsibility for its correct application is accepted by the end user.

### VAGINAL BIRTH AFTER CESAREAN

**Predicted chance of vaginal birth after cesarean:** 61.2%

95% confidence interval: [58.0%, 64.3%]

---

This calculator is based on the equation published in the article "Development of a nomogram for prediction of vaginal birth after cesarean" cited below. It is designed for educational use and is based on a population of women who received care at the hospitals within the MFMU Network. Responsibility for its correct application is accepted by the end user.


---

This calculator is based on the equation published in the article "Development of a nomogram for prediction of vaginal birth after cesarean" cited below. It is designed for educational use and is based on a population of women who received care at the hospitals within the MFMU Network. Responsibility for its correct application is accepted by the end user.

RACE IS A SOCIAL CONSTRUCT, RATHER THAN A BIOLOGICAL ONE.
RACE IS NOT A RISK FACTOR... RACISM IS.

LGBTQ IDENTITY IS NOT A RISK FACTOR... HOMOPHOBIA/TRANSPHOBIA IS.
DATA AS PROTEST.

DATA FOR BLACK LIVES
SOME NARRATIVES WERE CREATED BY DATA, AND CAN ONLY BE DISRUPTED BY DATA.

DATA IS A FORM OF PROTEST WHEN ALL OTHER CHANNELS OF ADVOCACY ARE BLOCKED.
DATA AS ACCOUNTABILITY.

DATA FOR BLACK LIVES
An Open Letter to Facebook from the Data for Black Lives Movement

Give Black researchers, data scientists and Black communities access to our data.

Over the past several years, Cambridge Analytica harvested the data of 87 million active Facebook users. We may never know what would have happened if this data did not land in the hands of individuals seeking to steer the country in the direction it is now. What we do know is that it’s not too late to do the right thing. This data, which would not exist without the 2.1 billion people who use Facebook, can and should be used to achieve a level of progress that fulfills the true promise of technology. Mr. Zuckerberg, I believe the world is ready for Facebook to live up to its mission—as “a tool for building community and bringing the world closer together.”

In the world we live in, data is destiny. For Black people, who have been disproportionately harmed by data-driven decision-making, this is especially true.

We urge you to work with Data for Black Lives and make a commitment to the following:

1. Commit anonymized Facebook data to a Public Data Trust.
2. Work with technologists, advocates, and ethicists to establish a Data Code of Ethics.
3. Hire Black data scientists and research scientists.
WASHINGTON, D.C. – U.S. Senator Cory Booker (D-NJ) today called on Facebook to enact a number of additional reforms on the heels of its recent announcement that it would undergo a civil rights audit and political bias review. Specifically, Booker pushed Facebook to use its data as a source for good by creating a public data trust and establishing industry-wide ethical frameworks.

"Announcing plans for a civil rights and safety audit is an important and significant step," Booker wrote in a letter sent today to Facebook Co-Founder and CEO Mark Zuckerberg. "However, the audit is one action among many Facebook should take to provide its users a more inclusive, equitable, and safe platform. I urge Facebook to partner with outside stakeholders to use Facebook’s vast trove of data as a force for positive change."

"I also strongly suggest that Facebook work with organizers, activists, technologists and data ethicists to establish
NUTRITION LABELS FOR ALGORITHMS

CAR INSURANCE
COLLEGE ADMISSION
GOOGLE SEARCH
FACEBOOK NEWSFEED
CHILD WELFARE
PREDICTIVE POLICING
FICO CREDIT SCORES
HEALTH INSURANCE
<table>
<thead>
<tr>
<th>Experian</th>
<th>Fair process</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fair outcome</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Privacy</td>
<td>F</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experian</th>
<th>Fair process</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fair outcome</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Clarity of what is predicted</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Solves problems for all</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Legitimate peer review</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Representation of all groups</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Adequate data and diverse outcomes</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>Privacy</td>
<td>F</td>
</tr>
</tbody>
</table>
DATA AS COLLECTIVE ACTION.
How community members in Ramsey County stopped a big-data plan from flagging students as at-risk

By Carrie Pomeroy | 5 hours ago
What are we optimizing?
Real HIV prevention is housing
real HIV prevention is ending mass incarceration
real HIV prevention is economic justice
real HIV prevention is ending school pushout
real HIV prevention is addressing racism
real HIV prevention is ending queerphobia
THANK YOU!

JOIN OUR MOVEMENT

@data4blacklives
@yeshican
yeshi@d4bl.org
15 MINUTE BREAK

INTRODUCE YOURSELF TO ANOTHER PERSON

Back at 1:45 CST

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
GROUP CHECK-IN

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
SHIFTING THE CONVERSATION TO HEALTH RESEARCH IN SEXUAL AND GENDER MINORITY POPULATIONS

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
SSOGI Data Capture: Gaps, Challenges, and Opportunities

Gregory Phillips II
Why Are Data on SGM Populations Important?

• Data are essential for identifying, examining, and working to improve disparities in sexual and gender minority (SGM) health

• Reliable sex, sexual orientation, and gender identity (SSOGI) measurements have the capability to assess how SSOGI variables relate to one another, impact collective life experiences, and influence health outcomes

However, due to poor and inconsistent data collection, there is an unknown number of SGM individuals being miscategorized or rendered invisible in federal health data.
Why Are Data on SGM Populations Important?

US LGBT Population by % – 2012-2017

Why Are Data on SGM Populations Important?

Source:
Why Are Data on SGM Populations Important?

- Family support
- Housing
- Health care access
- Education and economic security
- Victimization
- Mental health
- HIV
- Substance Use
- Diabetes
Why Are We Seeing SGM Health Disparities?

- Minority Stress
- Social Determinants of Health
- Stigma
- Anti-SGM Policies
Why Do We Need Better SGM Data?

• Understanding SGM disparities is an integral component in planning culturally competent interventions
  – This requires data that capture the full plurality of sexual and gender identity

• Data are essential in addressing disparities faced by individuals with intersectional identities including sexual orientation, gender identity, race, and ethnicity
  – Excluding questions capturing sexual orientation and inclusive gender identity perpetuates a lack of knowledge about SGM health
Capturing Sexual Orientation: the 3 Concepts

Sexual Orientation

Identity

Attraction

Behavior
SSOGI Definitions: Identity

• Identity
  – One’s internal sense of self
    • Identity terms are neither static nor definitive: meaning and connotation of terms can change over time (e.g., queer)
    • Identity can also fail to capture accurate information as to sexual behaviors and attractions

How do you describe your sexual orientation?
• Straight (Heterosexual)
• Gay/Lesbian
• Bisexual/Pansexual
• Asexual
• Not Listed: ______
• Prefer not to respond
• I don’t know how I describe my sexual orientation
• I don’t know what this question is asking
SSOGI Definitions: Attraction

• Attraction
  – A dimension of sexual orientation pertaining to the type of person one is attracted to, or a characterization of a lack of attraction
    • Measuring behavior alone fails to capture asexual individuals and individuals, particularly youth, who may not yet have explored feelings of sexual attraction

People are different in their sexual attraction to other people. Which best describes your feelings? Are you...
1. Only attracted to females
2. Mostly attracted to females
3. Equally attracted to females and males
4. Mostly attracted to males
5. Only attracted to males
6. Not sure
7. I do not experience sexual attraction
SSOGI Definitions: Behavior

• Behavior
  – The sex and sexual activities that an individual engages in or does not engage in
    • Typically measured by asking about the gender of one’s partners or sexual acts they engage in, but often cannot capture the full scope of an individual’s sexual experience
  – Identity, behavior, and attraction can change over time, and while someone may indicate a current sexual orientation, it may not capture the history of their sexual experience

During your life, with whom have you had sexual contact?
1. I have never had sexual contact
2. Females
3. Males
4. Females and males
SSOGI Measurement: Sexual Identity, Attraction, and Behavior

• Only measuring one construct of sexual orientation misses out on minority populations
  • E.g., bisexual and asexual identity

Identity

Bisexual

Attraction

Asexual

Behavior

No sex

All must be measured to capture required nuance and assess health risks.
How Do Sexual Identity and Behavior Intersect?
SSOGI Definitions: Sex

- Sex
  - What is your sex?
    - Male
    - Female

- Intersex individuals have sex characteristics that deviate from typical definitions of male and female bodies.

- Many transgender people legally and/or medically change their sex characteristics to be in alignment with their gender identity.
  - There is no single trans experience when it comes to sex.
SSOGI Definitions: Gender

Identity
• Your internal sense of self
• May or may not align with gender expression

Expression
• How you present to the world
• Includes things like pronouns, clothes, and more

The conflation of the constructs of sex, gender identity, and gender expression in many federal health surveys creates significant barriers to accurate and comprehensive measurement.
SSOGI Definitions: Transgender and Other Terms

**Transgender**
- Both an umbrella term and a specific identity
- Describes individuals whose gender identity does not align with the sex they were assigned at birth (SAB)
- Often associated with medical or social transition
- In older generations, you may hear “transsexual”

**Non-Binary**
- Both an umbrella term and a specific identity
- Describes individuals who do not identify as either male or female
- You may also hear: “Agender” or “Genderqueer”
- May or may not involve medical or social transition

**Gender Non-Conforming**
- Describes gender expression, but may also be a specific identity term for some
- Describes individuals whose gender expression does not align with their SAB or gender identity
<table>
<thead>
<tr>
<th>Measure</th>
<th>Constructs assessed</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey (Census Bureau)</td>
<td>Sex, sexual orientation (partner status)</td>
<td>Does not measure transgender-inclusive gender identity</td>
</tr>
<tr>
<td>Behavioral Risk Factor Surveillance System (BRFSS)</td>
<td>Sex, sexual orientation, gender identity (questions vary by state)</td>
<td>Limited sexual orientation options, doesn’t assess intersex status</td>
</tr>
<tr>
<td>Current Population Survey (Census Bureau)</td>
<td>Sex, sexual orientation (partner status)</td>
<td>Does not measure transgender-inclusive gender identity</td>
</tr>
<tr>
<td>Healthy Youth Survey</td>
<td>Sex, sexual orientation</td>
<td>Does not measure transgender-inclusive gender identity</td>
</tr>
<tr>
<td>Kaiser Permanente Member Health Survey</td>
<td>Sex, sexual orientation, transgender-inclusive gender identity</td>
<td>Does not measure asexual status or intersex status</td>
</tr>
<tr>
<td>Youth Risk Behavior Surveillance System (YRBSS)</td>
<td>Sex, sexual orientation, transgender status</td>
<td>Does not distinguish between sex or sex at birth, sexual orientation measures inconsistent, does not include intersex status</td>
</tr>
<tr>
<td>National Intimate Partner and Sexual Violence Survey</td>
<td>Sex, sexual orientation, transgender-inclusive gender identity</td>
<td>Does not measure intersex status</td>
</tr>
</tbody>
</table>
Effectiveness in Measuring SSOGI: Overall

- Only **19%** of data sources measure transgender-inclusive gender identity - most sources included only a single transgender-inclusive measure
- **40%** of sources measure a single dimension of sexual orientation
  - This can limit what can be known about sexual orientation and health
- Just over half of sources measure sexual identity with two or more dimensions of sexual orientation, and only **14%** follow best practice recommendations to measure such
- Only **one** survey assesses asexual identity
- While SSOGI data collection practices are improving, very few sources measure SSOGI consistently, across time, and according to best practices

Source: *Measuring Sexual and Gender Minority Populations in Health Surveillance*, DOI: [https://dx.doi.org/10.1089%2Flgbt.2016.0026](https://dx.doi.org/10.1089%2Flgbt.2016.0026)
Leveraging Population-Level Data: YRBS

• NIAAA funding to study HIV and alcohol use disparities between sexual minority and majority youth within YRBS

• Although this is a unique population-level dataset, there are multiple challenges:
  – YRBS has no distinction between current sex or sex at birth, measures sexual orientation inconsistently, measures transgender status as opposed to gender identity, and has limited sexual orientation options
  – **BUT**, it includes measures of all three SSOGI constructs
    • *This allows for the examination of intersecting (i.e., both sexual and gender minority) minority status*
Some YRBS Papers Our Team Has Published


Despite imperfect data collection on transgender individuals, YRBS has led to groundbreaking work

- From the CDC: *Transgender Identity and Experiences of Violence Victimization, Substance Use, Suicide Risk, and Sexual Risk Behaviors Among High School Students — 19 States and Large Urban School Districts, 2017*
SSOGI-Inclusive Survey Landscape: Conclusion and Implications

• Data systems often miss sexual orientation and gender identity
  – Even those that include questions regarding SSOGI often cannot gauge
    the full scope of identities
    • Individuals and communities are left out in the process
    • Failure to oversample SSOGI minorities results in low power to detect effects

• Nuance and intersectional identities are lost when data is
  collapsed and grouped together
  – e.g., grouping together “gay/lesbian/bisexual” as an option for sexual
    orientation not only removes the individual identities of gay, lesbian, and
    bisexual individuals, but leaves out asexual individuals altogether
    • Failure to capture the full scope of gender identity, legal sex, and current sex further
      erases nuanced identities and paints an inaccurate picture of SGM populations
SHIFTING THE CONVERSATION TO HEALTH RESEARCH IN SEXUAL AND GENDER MINORITY POPULATIONS

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
Ethical Considerations for Conducting Big Data Research in LGBTQIA Populations

Lauren Beach
@lauenbbeach
Refresher: How to join the conversation remotely

• BlueJeans
  – Write a comment in the public chat
  – Send a private message to the moderator

• Twitter
  – Follow us @CONNECT_ISGMH, we’ll follow back!
    • This allows you to Direct Message (“DM”) our team questions
  – Use the #BiasInBigData hashtag when responding in real time

• Anyone can email us a question or comment at connect@northwestern.edu
Scholarly articles for **LGBT big data**

- Public attention: Agenda-setting theory meets big data - Russell Neuman - Cited by 321
- Making big data, in theory - Boellstorff - Cited by 166
- Out smoking on the big screen: tobacco use in LGBT... - Lee - Cited by 20

People also ask

- What is the gayest city?
- What is the gayest country in the world?
- What is the statistics of homosexuality?
- What percentage of LGBT youth are homeless?
Other than survey data, what other sources of SGM big data are there?
What sources of big data do you use in your analyses?
SSOGI Big Data Sources

• Surveillance data
• Electronic medical records
• Medical billing datasets
• Social media data
• Market data
• Program evaluation data
• Government databases
Data Integration as a SGM Big Data Research Tool

- Linking Identifier
- Linking Identifier
- Linking Identifier
Data Integration as a SGM Big Data Research Tool
Data Integration as a SGM Big Data Research Tool: The Rainbow Connection
Data Integration as a SGM Big Data Research Tool: The Rainbow Connection
The Ethics of Data Integration in SGM Big Data Analyses
“Good” Example:
Data Integration Can Facilitate Important SGM Health Studies That Are Not Otherwise Feasible
Data Integration as a SGM Research Tool

NCHS Data Linkage

NCHS Data Linked to NDI Mortality Files

NCHS has linked various surveys with death certificate records from the National Death Index (NDI). Linkage of the NCHS survey participants with the NDI mortality data provides the opportunity to conduct a vast array of outcome studies designed to investigate the association of a wide variety of health factors with mortality. The restricted-use Linked Mortality File (LMF) has been updated with mortality follow-up data through December 31, 2015.

NCHS Surveys Linked to NDI Mortality Data

- National Health Interview Survey (NHIS): 1985 - 2014
- Third National Health and Nutrition Examination Survey (NHANES III)
- Second National Health and Nutrition Examination Survey (NHANES II)
- NHANES I Epidemiologic Follow-up Study (NHEFS)
- National Home and Hospice Care Survey (NHHCS) 2007
- Supplement on Aging (SOA)
- The Second Longitudinal Study of Aging (LSOA II)
Bad Example:
SGM Data Integration Can Lead to Context Collapse – and Outing
Being “Outed” Can Have Serious Consequences for LGBTQIA People

• LGBTQIA people are systematically socially stigmatized and marginalized
  – LGBTQIA people may experience rejection from family and friends
  – There are no federal anti-discrimination protections for LGBT people
LGBTQIA People May Intentionally Curate “Data Siloes” to Manage SGM Identity Disclosure

“Too Gay for Facebook”: Presenting LGBTQ+ Identity Throughout the Personal Social Media Ecosystem

MICHAEL A. DEVITO, Media, Technology, and Society, Northwestern University, USA
ASHLEY MARIE WALKER, Media, Technology, and Society, Northwestern University, USA
JEREMY BIRNHOLTZ, Communication Studies, Northwestern University, USA

Most US social media users engage regularly with multiple platforms. For LGBTQ+ people, this means making self-presentation decisions not just on one platform, but many. These choices are made in the face of sometimes-overlapping platform environments, which can have consequentially different norms, audiences, and affordances. Moreover, many LGBTQ+ users face high stakes in online self-presentation, due to the risk of stigmatization of their LGBTQ+ identity, increasing the importance of self-presentation decisions that enable them to achieve their goals and avoid stigmatization. This combination of environmental complexity and high stakes is not adequately accounted for in existing work on self-presentation, but doing so is important to support and understand the experiences of LGBTQ+ and other potentially stigmatized users. We adopt an ecological approach to an interview and cognitive mapping study of 20 LGBTQ+ social media users. We find that participants employ the platforms, audiences, affordances, and norms within what we call their “personal social media ecosystems” to avoid stigmatization while still allowing for expression of their LGBTQ+ identity and the flexibility to adjust their presentation over time.

CCS Concepts: • Human-centered computing: Empirical studies in collaborative and social computing • Human-centered computing: Empirical studies in HCI • Social and professional topics: Sexual orientation

KEYWORDS
Sexual and gender minorities; LGBTQ; LGTB; queer; self-presentation; social media; technology ecosystems; identity management; disclosure; privacy

Northwestern

Institute for Sexual and Gender Minority Health and Wellbeing
What about Algorithms?
Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

Contributors: Yilun Wang, Michal Kosinski
Date created: 2017-02-15 10:37 AM | Last Updated: 2018-10-23 04:05 PM
Category: Project

Description: We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. Human judges achieved much lower accuracy: 61% for men and 54% for women. The accuracy of the algorithm increased to 91% and 83%, respectively, given five facial images per person. Facial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles. Prediction models aimed at gender alone allowed for detecting gay males with 57% accuracy and gay females with 58% accuracy. Those findings advance our understanding of the origins of sexual orientation and the limits of human perception. Additionally, given that companies and governments are increasingly using computer vision algorithms to detect people's intimate traits, our findings expose a threat to the privacy and safety of gay men and women.

How many people have heard of this paper?
Do you think it should have been published?
What would you have done differently?
Dating Profiles are Like Gay Bars: Peer Review, Ethics and LGBTQ Big Data

Posted on September 13, 2017

(Posted 9/13; updated 9/14. My original critique is here; and my resource of relevant blogs, media reports, and Kosinski statements is here).

“Under ethical review,” announced The Outline, of the Wang and Kosinski pre-print paper that controversially claimed to use face-recognition technology to detect intrinsic differences between homosexual and heterosexual faces. The statement came from the editor of the Journal of Personality and Social Psychology (JPSP), which peer-reviewed and accepted the paper. This smacks of the journal throwing Wang and Kosinski under the bus [UPDATE: Hal Hodson reports that the publication will go ahead, as it should].

This post explains why the journal bears the brunt of the blame here, and how this controversy is symptomatic of two larger problems: the crisis in academic peer review, and the general disregard of academic expertise on sexuality and LGBTQ lives. The LGBTQ community has long had concerns about research, privacy, and consent: we don’t treat gay dating profiles or gay bars as public places, and neither should researchers.

WARNING: Any institution using this or any of its associated sites for study or projects – you DO NOT have my permission to use any of my profile or pictures in any form or forum both current or future. If you have or do, it will be considered a serious violation of my privacy and will be subject to legal ramifications.

Potential Solutions

Engage SGM people, community organizations, and researchers in the design, conduct, analysis, and dissemination of SGM Big Data Analyses

“Nothing about us without us”
Potential Solutions

Enhance opportunities for teaching, training, and mentorship in SGM research
Potential Solutions

Develop ethical guidelines and recommendations for conducting (SGM) big data analyses
Group Exercise:
Brainstorming SGM Big Data
Ethical Recommendations
Shaping Principles & Frameworks
Data Integration
Data Privacy
Data Security
Community Engagement
Ethical Review Standards
HOW TO PROMOTE THE INCLUSION OF DIVERSE VOICES IN BIG DATA AND DATA SCIENCE?

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
Individual and structural factors predicting HIV testing among Latinx MSM: Substance use as a moderator

Austin C. Eklund, Ed.M., M.A., Frank R. Dillon, Ph.D., and Ryan C. Ebersole, Ph.D.

AIDS Care (2019)
Motivation for the study

- Used data from an NIMHD grant-funded longitudinal study investigating Latinx MSM and their HIV-related health behaviors
  - PI: Frank R. Dillon, Ph.D. (co-author)

- Associating more substance use with fewer health-promoting behaviors, as is commonly held, ignores historical context of queer life

  - Queer communities have taken refuge in bars, clubs, and other social settings where they can be with similar folks and where alcohol and drugs may be consumed

  We therefore wanted to see how substance use actually impacts HIV testing behaviors by examining interactions between individual and structural factors, some of which were specific to LMSM (e.g., machismo, caballerismo)
Some interesting findings

- Only 21% of the sample reported having been tested for HIV in the previous 12 months ($n = 502$)
  - Explicable due to sampling frame: not just clinics, but bars, clubs, bathhouses, etc.—across identity spectrums

- More HIV testing was reported among:
  1. LMSM who use more drugs and who report more exposure to prevention programming (2.5x higher odds)
  2. LMSM who use more drugs and who report more sexual identity commitment (6.5x higher odds)
  3. LMSM who use less alcohol and who self-identify as more homosexual (Kinsey scale; .25x higher odds)
  4. LMSM who use more alcohol and who report more access to health care (2.33x higher odds)
HOW TO PROMOTE THE INCLUSION OF DIVERSE VOICES IN BIG DATA AND DATA SCIENCE?

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData
WHERE DO WE GO FROM HERE?

We’re also livestreaming @CONNECT_ISGMH
Keep the conversation connected with #BiasInBigData