Big Data
*For or Against*
Health Disparities

Deborah Guadalupe Duran, PhD
Director, Office of Science Policy, Planning Reporting and Data
National Institute of Minority Health and Health Disparities
OVERVIEW

“Imagine most of the world is getting healthier because of some new technology, but you’re getting left behind”

• NIMHD Mission

• Need for Big Data – Minority Health/Health Disparities

• Big Data: Structured and Unstructured (Use with Caution)

• Big Data: Importance for Health Disparities Research and Clinical Care

• NIMHD - Big Data into Health Disparities Research
NIMHD Mission
NIMHD’S MISSION

• To improve minority health
• To reduce health disparities

To achieve:
• Plans, coordinates, reviews, and evaluates NIH minority health and health disparities research and activities
• Conducts and supports research in minority health and health disparities
• Promotes and supports the training of a diverse research workforce
• Translates and disseminates research information
• Fosters innovative collaborations and partnerships
MINORITY AND *HEALTH DISPARITIES POPULATIONS

- Racial/Ethnic Minorities
  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Hispanic/Latino American
  - Native Hawaiian or Other Pacific Islander

- *Low socio-economic status
- *Rural
- *Sexual and gender minorities
- *Others subject to discrimination who have poorer health outcomes

Source: Duran DG, Perez-Stable, EJ, Novel Approaches to Advance Minority Health and Health Disparities Research, AJPH Supplement 1, 2019
“Health characteristics and attributes of racial and/or ethnic minority groups who are socially disadvantaged due in part by being subject to potential discriminatory acts.”

Minority health research examines singularly and in combination the attributes, characteristics, behaviors, biology, and other factors that influence the health outcomes of minority racial and ethnic groups, including within-group or ethnic subpopulations, with the goal of understanding mechanisms and improving health.
HEALTH DISPARITIES DEFINITION

A health difference that adversely affects defined disadvantaged populations, based on one or more health outcomes.

HD OUTCOMES:

• Higher incidence/prevalence, including earlier onset or more aggressive progression
• Premature or excessive mortality
• Greater global burden
• Worse self-reported outcomes measures that reflect daily functioning or symptoms from specific conditions
• Poorer health behaviors and clinical outcomes (related to above)

Source: Duran DG, Perez-Stable, EJ, Novel Approaches to Advance Minority Health and Health Disparities Research, AJPH Supplement 1, 2019
## Causes

### Health Determinants
- Behaviors
- Biology
- Sociocultural
- Physical Environment
- Health Care & Systems

*These influence health disparities*

### Biological
- Biochemical
- Genome and epigenome
- Proteome
- Microbiome
- In utero exposure
- Metabolic factors
- Pathogens
- Physiologic responses to stress/Allostatic load
- Organ systems
- Nervous system
- Telomere/Cellular aging/Genes
- Cellular functions & communication
- Enzymes
- Inflammation
- Demographics
- Endocrine system/Hormones

## Effects

### Health Disparity Outcomes
- Higher incidence prevalence, including earlier onset or more aggressive progression
- Premature or excessive mortality from specific conditions
- Greater global burden as indicated by population health measures
- Poorer health behaviors and clinical outcomes (related to above)
- Worse self-reported outcomes measures that reflect daily functioning or symptoms from specific conditions

## Health Determinants

### Behavioral
- Tobacco use
- Diet and nutrition
- Preventive health behaviors
- Unprotected sexual intercourse
- Domestic/Family violence
- Physical activity
- Substance use, abuse, misuse, and addiction
- Compliance and adherence with prescribed therapy
- Delays in seeking care after symptom awareness
- Living responsibly with infectious disease
- Hygiene/Oral hygiene
- Cultural beliefs/Schemas
- Religious beliefs/Schemas

### Biological
- Biochemical
- Genome and epigenome
- Proteome
- Microbiome
- In utero exposure
- Metabolic factors
- Pathogens
- Physiologic responses to stress/Allostatic load
- Organ systems
- Nervous system
- Telomere/Cellular aging/Genes
- Cellular functions & communication
- Enzymes
- Inflammation
- Demographics
- Endocrine system/Hormones

### Sociocultural Environment
- Employment status and security
- Income
- Housing and food security
- Health insurance status (affordability/quality)
- Social and economic adversity and/or inequality
- Immigration and legal status
- Geographic location
- Residential segregation
- Educational attainment
- Access to quality education
- Transportation options
- Limited English Proficiency
- Health literacy/numeracy
- Discrimination, racism, and stigma
- Health socialization and education
- Psychosocial stressors
- Historical trauma
- Social safety net
- Community reentry (e.g., prison, military service)

### Physical Environment
- Housing status
- Neighborhood violence
- Unhealthy housing units
- Residence crowding
- Exposure to toxic substances (e.g., pollution, radiation, lead, mold, dust mites)
- Aesthetic elements (e.g., trees)
- Access to safe recreational facilities
- Quality of air and water
- Concentration of fast-food outlets & access to full-service grocery stores
- Public safety (e.g., fire dept., police)
- Occupational conditions and/or hazards
- Affordability of resources

### Clinical Events & Health Care Systems
- Patient-clinician communication
- Health insurance coverage/policies
- Access to preventive services and/or quality health care
- Disease management & functioning status
- Symptom and pain management
- Drug interactions and synergies
- Use of alternative therapies
- Appropriate diagnostics
- Access to emerging technologies
- Access to public health education, information, and health alerts
- Precision medicine
- Generalizability of research findings
- Translation of research
- Dissemination & diffusion of research results
- Macro-structural stressors (e.g., policies & procedures)
- Incorporation of spiritual and/or traditional healers
- Institutional discrimination in health care
- Health care system mistrust
- Culturally competent care
- Workforce diversity
- Electronic medical records
- Palliative and end-of-life care
- Living with chronic illness and/or comorbid conditions
- Long-term care
- Access to health information/consent in primary language
- Policies & political practices
- Diversity of biomedical/health delivery workforce
Our environments cultivate our, communities and our communities nurture our health.

When inequities are high and community assets are low, health outcomes are worst.

Violence
Substance Abuse
HIV/AIDS
Infant Mortality
Malnutrition
Stress
Obesity
Depression
Heart Disease

When inequities are low and community assets are high, health outcomes are best.

HIV/AIDS
Infant Mortality
Heart Disease
Malnutrition
Stress
Depression
Substance Abuse
Smoking
Violence
Obesity

Sense of Community
Social Networks
Social Support
Participation
Leadership
Political Influence
Organizational Networks

Quality Schools
Access to Healthy Foods
Access to Healthcare
Clean Environment
Transportation Resources
Adequate Income
Health Insurance
Quality Housing
Jobs

Adverse Living Conditions
Poverty
Segregation
Marketing for Tobacco and Alcohol
Entrepreneurial Economies
Environmental Toxins
Unemployment
Discrimination

Fragmented Systems
Restricted Power
Disinvestment
Disconnected Members

From Brennan Ramirez et al, *Promoting Health Equity*, CDC, 2008 et al. Figure adapted from Anderson et al, 2003; Marmot et al, 1999; and Wilkinson et al, 2003

NIMHD INVESTED IN BIG DATA

• Need Data
  o Population representation that is adequate and accurate
  o Social determinants of health, the causes of health disparities

• Improve
  o Clinical care
  o Population health/Reduce health disparities

• Prevent additional health disparities generated by technological advances
Need for BIG DATA

TO IMPROVE MINORITY HEALTH
TO REDUCE HEALTH DISPARITIES
DEMOGRAPHICS – USA 2017 to 2019 Data

- Estimated population of 328,285,992 (~40% minorities/~50% HD populations)
- 82.3% of the population resides in cities and suburbs
- Hispanic and Latino Americans 18.1%
- African Americans 13.4%
- Asian Americans 5.8% / fast growing: 3% growth rate
- American Indian/Alaska Native Native 1.3%
- SGM/LGBT: 3-5% or 9 million Americans (adult population)
  - Estimated 19 million Americans (8.2%) report that they have engaged in same-sex sexual behavior
  - 25.6 million Americans (11%) acknowledge at least some same-sex sexual attraction.
- Poverty: 13.5% or 39.7 million Americans (2017 data)
- Rural: 19.3% or 60 million Americans (2017)

https://www.census.gov/quickfacts/fact/table/US/PST045218
https://poverty.ucdavis.edu/faq/what-current-poverty-rate-united-states
Asians Projected to Become the Largest Immigrant Group, Surpassing Hispanics

FIGURE 5

Asians Projected to Become the Largest Immigrant Group, Surpassing Hispanics

% of immigrant population

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual</th>
<th>Projected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

Note: Whites, blacks, and Asians include only single-race non-Hispanics. Asians include Pacific Islanders. Hispanics are of any race. Other races shown but not labeled.

Source: Pew Research Center estimates for 1965-2015 based on adjusted census data; Pew Research Center projections for 2025-2065

PEW RESEARCH CENTER
Hispanic Immigration Declining

Immigrant share falls among largest Hispanic origin groups since 2000

% foreign born of each origin group

Note: “Immigrants” includes those born outside the U.S. or its territories (including Puerto Rico) to non-U.S. citizen parents. People in group quarters such as college dormitories or institutions are not included in figures for 2001 to 2005. Due to changes in the wording of the Hispanic origin question in the 2000 census some Hispanic origin groups may have led to many not indicating their Hispanic origin, resulting in low population estimates. For more see http://www.pewhispanic.org/2002/05/09/counting-the-other-hispanics/

Source: Pew Research Center tabulations of 2000 census (5% IPUMS) and 2001-2015 American Community Surveys (1% IPUMS).

PEW RESEARCH CENTER
Americans are more Racially and Ethnically Diverse than in the past, and the U.S. is projected to be even more diverse by 2055.
### INCREASING LOWER CLASS

<table>
<thead>
<tr>
<th>Year</th>
<th>Lower</th>
<th>Lower Middle</th>
<th>Middle</th>
<th>Upper Middle</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>20</td>
<td>9</td>
<td>50</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>2011</td>
<td>20</td>
<td>9</td>
<td>51</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>2001</td>
<td>18</td>
<td>9</td>
<td>54</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>1991</td>
<td>18</td>
<td>9</td>
<td>56</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>1981</td>
<td>17</td>
<td>9</td>
<td>59</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>1971</td>
<td>16</td>
<td>9</td>
<td>61</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

Note: Adults are assigned to income tiers based on their size-adjusted household income in the calendar year prior to the survey year. Figures may not add to 100% due to rounding.


PEW RESEARCH CENTER
USA EDUCATION

- US Overall High School Education 89%
- College Degree
  - Whites 37%
  - Blacks 23%
  - Hispanics 16.4%
  - Asian American 55.9%
- Students from low and moderate income households could afford to pay for just 1% to 5% of college tuition in USA
- Adults with HS education earned - $35,615 ave
- Adults with BS degree earned - $65,482 ave

Census Data 2016
1950-2010 POPULATION: AGING
USA DOUBLED
GLOBAL TRIPLED

Estimates of the Global Population, by Age, 1950 to 2050

PEW RESEARCH CENTER
Structured
Unstructured

Use with Caution
Platforms

• **Structured** - organized such that data can be described and analyzed in conventional ways, including hypothesis testing. The unit is an individual with well-defined characteristics.

  *Examples: Medicare, SEER-Medicare, National Health Interview Survey, American Community Survey*

• **Unstructured** – analyzed using a new discipline, data science that looks for patterns. The unit of analysis is usually a location without well-defined personal characteristics. Artificial intelligence algorithms.

  *Examples: Genome data, Google searches, Facebook, mobile sensors, and data collected for marketing purposes*
Two Ways of Knowing: **Structured** and **Unstructured Data**

Source: Sim I., Two Ways of Knowing: Big Data and Evidence-Based Medicine
Annals of Internal Medicine 2016
Race/Ethnicity in Structured Data

- Census data has evolved
  - Beginning in 1960, Americans reported their own race
  - Beginning in 1980, ancestry/ethnic origin was asked
  - Since 2000, respondents could report more than one race
  - Missing racial categories
  - Hispanic/Latino (ethnicity)

- Census data are used to:
  - Make policy and allocate funding ($800 billion)
  - Test hypotheses

- Gold Standard: Self Reported Race/Ethnicity

- Ancestry biomarkers

- Movement to collect Social Determinants of Health
Unstructured Data: Adequate and Accurate Racial/Ethnic Data?
Genome / Omic Data Bases
Misclassified Ancestry in AA Cancer Cell Lines  (inadequate and inaccurate)

• Cell lines are used by cancer researchers to probe molecular mechanisms. Few cell lines are non-European

• Observed differences in how certain cancers behave at the biological level in various ethnic groups, highlighting the need for ethnically diverse cell lines to probe the molecular basis for these differences.

• A team of researchers analyzed genetic ancestry of 15 prostate, breast, and cervical cancer cell lines and found that several lines labeled as mixed or black/African-American were misclassified
  o A common prostate cancer cell line classified as African-American actually carries more than 90% European ancestry (22Rv1 had 99% European ancestry)
  o E006AA-hT, (prostate) labeled as African-American actually carries 91% European ancestry.
  o HeLa (Heneritta Lacks) cells contain 66% West African ancestry, which is slightly lower than the average of about 80% West African ancestry for US born African-Americans.

• Published studies using misclassified cell lines may be inadequate to understand the disease within a population group and inaccurate for the development of personalized medicines

2016 meta-analysis: 2,511 studies from around the world found that 81% of participants in genome-mapping studies were of European descent. (2009, 96% of European descent)
‘Omic Limited Data Sources

• International Cancer Genome Consortium (ICGC)-20 countries
  32 projects in USA, 1 in India and 0 in Africa
• Some inherent bias in genomic data stems from non-scientific roots
  o Over-represent people who get sick more
  o Demographics around hospital
  o Clinical trial enrollment
  o Costs
• 2015 study found that 2% of the more than 10,000 NIH-funded cancer studies include adequate sample of minority groups to measure statistically significant change.
• Too frequently, rely on observation classification for race/ethnicity

https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks/
Artificial Intelligence
Artificial Intelligence (AI) Plays Integral Role in Medical Diagnostics

• AI will sift through the vast amounts of health information produced by our society
  o Find patterns that will help us be more efficient, wealthier and happier
  o Assist clinical decision making
  o Reduce diagnostic and therapeutic errors
  o Inferences for health risks

• To address diagnostic error, researchers and companies are leveraging artificial intelligence
  o **Chatbots**: Speech recognition capability to identify patterns in patient symptoms to form a potential diagnosis, prevent disease and/or recommend an appropriate course of action.
  o **Oncology**: Train algorithms to recognize cancerous tissue at a level comparable to trained physicians.
  o **Pathology**: Pathology is the medical specialty that is concerned with the diagnosis of disease based on the laboratory analysis of bodily fluids such as blood and urine, as well as tissues. Machine vision and other machine learning technologies can enhance the efforts traditionally left only to pathologists with microscopes.
  o **Rare Diseases**: Facial recognition software is being combined with machine learning to help clinicians diagnose rare diseases. Patient photos are analyzed using facial analysis and deep learning to detect phenotypes that correlate with rare genetic diseases.

Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)

- US justice system, reviled for its racial bias, turned to technology for help-algorithms had a racial bias too.

  - More prone to mistakenly label black defendants as likely to reoffend – wrongly flagging them at almost twice the rate as white people (45% to 24%)

- Black defendants were 77 percent more likely to be assigned higher risk scores than white defendants.

- White defendants who re-offended within the next two years were mistakenly labeled low risk almost twice as often as black re-offenders (48 percent vs. 28 percent)

- PedPol Crime Hot Spots: Suggested majority black neighborhoods at about twice the rate of white ones / when the statisticians modelled the city’s likely overall drug use, based on national statistics, it was much more evenly distributed.

https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/
Dark-Skinned People & Females

• AI-Driven Dermatology: potential to save thousands of people from skin cancer each year—while putting others at greater risk
  o African-Americans have the highest mortality rate for skin cancer so AI trainers are missing skin cancer in African-Americans
  o Particularly bad at identifying women of color: 34% less accurate at recognizing darker-skinned females compared to lighter-skinned males

• Three of the latest gender-recognition AIs (Megvii)
  o Id person’s gender 99% of time for white men
  o Id person’s gender 35% of time for dark-skinned women

• Google search for CEO
  o 11% resulted in women (27% female CEOs in US)
  o Online advertising showed high-income jobs to men more than women

https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/
Discriminating Algorithms
Amplified Racist Sexist Age Biases

• If you don’t teach the algorithm with a diverse set of images, then that algorithm won’t work out in the public that is diverse. Algorithms have:
  o mistaken images of black people for gorillas
  o misunderstood Asians to be blinking when they weren’t
  o “judged” only white people to be attractive
  o Linkedin showed a preference for male name searches
  o Chatbot-Tay learned from twitter and began spouting anti-semitic messages

• AI systems are calibrated for younger, more urban bodies because poorer and older members of society don’t have access to the digital technologies and therefore do not end up in trainer data

• The highly selective nature of randomized control trials (RCT) systemically disfavors women, racial/ethnic minorities, the elderly, and those with co-morbidities

https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/
Weapons of Math Destruction

- Public/Patient perception might be that the algorithms are impartial.
- Garbage In-Garbage Out / Racism In-Racism Out (Khan 2017)
- Danger of outsourcing decisions to computers with biases
- Most vulnerable in society who are exposed to evaluation by automated systems
- Prevent AI from amplifying the inequalities of our past and affecting the most vulnerable members of our society. Data fed the machines reflects the history of our own unequal society--in effect, asking the program to learn our own biases.
- Bryson describes the way that machine learning can expose the prejudices embedded in our use of language. It encodes both the wisdom and the folly of all those who have used language.

Guardian Inequality Project: Rise of the racist robots – how AI is learning all our worst impulses by Stephen Buranyi Aug 2017
Do Not Automate Biases
Computers Learn from Us

• The formula for AI-powered systems is almost the same across all medical disciplines
  o Gather data on patients
  o Use it to predict what will happen when a new patient comes in

• AI Training data are crucial to accuracy.
  o If data don’t accurately represent a population of patients then any algorithm relying on the data is at a higher risk of making a mistake
  o Patients not in the training data set will not get an accurate diagnosis

• Deep learning – finding patterns in the data it is trained on – is especially susceptible to bias

The Food and Drug Administration (FDA) is now paying close attention to subject demographics in AI software.
AI Algorithms Need to Include Groups at Risk of Health Disparities

- Applications of AI in medical diagnostics are in the early adoption phase with limited data on patient outcomes.

- AI applications have the potential to impact:
  - How clinicians and health care systems approach diagnostics.
  - How individuals understand changes to their health in real-time.

- Rigorous testing of applications will be needed to validate utility.

Unless AI algorithms represent all racial/ethnic and other groups at risk of health disparities, undetected disease will result in **undetected disparities**.
Consider what could happen when doctors begin relying on AI to diagnose diseases like skin cancer, or determine which drug treatment is best for a serious illness based on biological markers. In products like Google Photos, these biases reaffirm false stereotypes and have detrimental impacts on users, but in the field of medicine, they can be the difference between life and death.

Bias at any point in data handling for precision medicine can lead to the recapitulation of longstanding health disparities,” wrote Ferryman and Pitcan in a February report for Data & Society. That means a world where black Americans die more often than white Americans will, at best, remain unchanged; worse, that demographic gap could widen.

In health care, these biased outcomes tend to mean one group of people gets better medical treatment than another. Groups are often characterized by gender or race, or other traits, like language, skin type, genealogy, or lifestyle.
Successful AI Kidney Matching

- AI can identify potential donors and recipients who are biologically suited for one another; but not who gets one
- The algorithms evaluate all the transplants possible among the patient-donor pool at once. Matches are made primarily on biological suitability, with the hardest-to-match patients getting first priority. The technology weighs criteria including the time the recipient has been on the waiting list, his or her age (children get priority), and whether the person who needs a kidney has been a living organ donor in the past (with the reasoning that people who have stepped up to give before should get priority if they one day find themselves in need).
- Machine making a decision based on what it has learned about human values.

In the kidney question, fair-minded principle that kidneys should be given to the people who will likely have the most years of productive life after receiving one. Before a computer can calculate the longevity of potential recipients, scientists have to feed the algorithm data on life expectancy for various populations.

But this leads to some problems. Men tend to die before women do. Black Americans die younger than Americans of any other ethnicity. A 65-year-old white woman in the US could expect to live another 20.5 years in 2015, four years longer than a black man of the same age.

What started with good intentions ends in systematic racial and gender discrimination.

Back to the drawing board – deal with it in beginning
Importance of Big Data for Health Disparities Research and Clinical Care for Minorities

FUSING
Fusing Data Platforms

- **Structured** big data (surveys, health/medical records, and administrative data) have the advantage of including demographic information about individuals included, and a defined population with known representation
  - Disadvantage is lack of granularity.
    - National Health Interview Survey is designed to measure outcomes at the national level
    - Behavioral Risk Factor Surveillance system is designed to measure outcomes at the state level

- **Unstructured** big data (Google, Facebook) has the advantage of millions of observations, hence plenty of granularity
  - Disadvantage is lack of knowledge about the representativeness of the data.
    - Who is included?
    - Who is missed?

- Limitations suggest that unstructured data needs to be fused with structured data for maximum utility in health disparities research

Source: Sim I., Two Ways of Knowing: Big Data and Evidence-Based Medicine, Annals of Internal Medicine 2016
Fusing Data: Mining for Racism

- The study calculated an Internet-based geographic measure of “racism” from Google searches of 196 market areas.
  
  **More N-word hits = More racism**

- Used geographic area racism measure to test for an association with higher African American mortality (structured data) in those same areas

- Found a one standard deviation increase in area racism was associated with a 6% increase in the all-cause Black mortality rate, after adjusting for covariates

  **More racism = Excessive AA Mortality Rates**

Fusing Data: Environmental Health Disparities

When it comes to Your Health, LOCATION MATTERS

Massachusetts study found that two air pollutants decreased significantly between 2003–2010, but racial and ethnic disparities in exposure remained

Fused Environmental Data with Census Data and Mortality Data
Relative Impact of Determinants on Health – Need Social Economic Data

- Clinical Care: 20%
- Health Behaviors: 30%
- Physical Environment: 10%
- Social & Economic Factors: 40%

Clinicians Need to Know about their Patients’ Social Determinants

*The Spirit Level* by Wilkinson and Pickett provided evidence that decreasing income, education, social status, and social support is correlated with increased morbidity and premature mortality.

A survey conducted by the Robert Wood Johnson Foundation found that 4 out of 5 physicians do not feel confident in their capacity to meet their patients’ social needs, and they believe this impedes their ability to provide quality care.

Examples of social determinants role in clinical care:

- Treatment options – social risk
- Adherence
- Better care for underserved marginalized individuals (reduce health disparities)
- Impact on DNA (other biological changes)
Who you are as a person is not just defined by your DNA, but by which parts of it your epigenome permits to be expressed—

Fusing Poverty Stressors & DNA

- **Stresses of being poor have a biological effect that can last a lifetime.** Lower socioeconomic status during adolescence is associated with an increase in methylation of the proximal promoter of the serotonin transporter gene,” which primes the amygdala—the brain’s center for emotion and fear—for “threat-related amygdala reactivity.” While there may be some advantages to being primed to experience high levels of stress (learning under stress, for example, may be accelerated), the basic message of these studies: Chronic stress and uncertainty during childhood makes stress more difficult to deal with as an adult.

- **Evidence suggesting that these effects may be inheritable,** whether it is through impact on the fetus, epigenetic effects, cell subtype effects, or something else. The science of the biological effects of the stresses of poverty, although in its early stages, has presented us with multiple mechanisms through which such effects could happen, and many of these admit an inheritable component. If a pregnant woman, for example, is exposed to the stresses of poverty, her fetus and that fetus’ gametes can both be affected, extending the effects of poverty to at least her grandchildren. And it could go further.

- **Scientific understanding of the experience of poverty can also inform medical treatments later in life.** Hippocampus samples from suicide victims with a history of childhood abuse and tested for DNA methylation controlling the expression of the gene NR3C1. They discovered an increased methylation around the NR3C1 promoter, which, in other studies, has been directly linked to a reduced expression of a protein called brain-derived neurotropic factor (BDNF). BDNF is among the most active neurotrophic factors, which drive the growth and development of new neurons even in adulthood. And the degree to which it is expressed may be inheritable. A 2015 study linked NR3C1 and reduced expression of BDNF in infants born to mothers who reported prenatal depressive symptoms.

NIMHD: Big Data into Health Disparities Research
Data Science Diversity Gap-Students

Race/Ethnicity of General Assembly Students by Course

<table>
<thead>
<tr>
<th>Course</th>
<th>White</th>
<th>Asian (S/E/SE)</th>
<th>Hispanic or Latino</th>
<th>African-American</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>46.9%</td>
<td>22%</td>
<td>9.8%</td>
<td>7.7%</td>
<td></td>
</tr>
<tr>
<td>Product Management</td>
<td>50.5%</td>
<td>21.4%</td>
<td>10.1%</td>
<td>5.7%</td>
<td></td>
</tr>
<tr>
<td>Digital Marketing</td>
<td>49.8%</td>
<td>15.9%</td>
<td>12.2%</td>
<td>8.1%</td>
<td></td>
</tr>
<tr>
<td>Data Science</td>
<td>46.1%</td>
<td>28%</td>
<td>7.8%</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>JavaScript Development</td>
<td>46.1%</td>
<td>19.7%</td>
<td>9.5%</td>
<td>7.7%</td>
<td></td>
</tr>
<tr>
<td>Front-End Web Development</td>
<td>45.5%</td>
<td>20.8%</td>
<td>11.0%</td>
<td>10.9%</td>
<td></td>
</tr>
<tr>
<td>User Experience Design</td>
<td>44.4%</td>
<td>24.5%</td>
<td>9.0%</td>
<td>6.9%</td>
<td></td>
</tr>
<tr>
<td>Data Analytics</td>
<td>43.8%</td>
<td>24.6%</td>
<td>9.3%</td>
<td>9.8%</td>
<td></td>
</tr>
</tbody>
</table>

Percentage of Hispanic/Latino and African-American Students Enrolled in Part-Time GA Courses

<table>
<thead>
<tr>
<th>Course</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>17.5%</td>
</tr>
<tr>
<td>Front-End Web Development</td>
<td>21.9%</td>
</tr>
<tr>
<td>Digital Marketing</td>
<td>20.3%</td>
</tr>
<tr>
<td>Data Analytics</td>
<td>19.1%</td>
</tr>
<tr>
<td>JavaScript Development</td>
<td>17.2%</td>
</tr>
<tr>
<td>User Experience Design</td>
<td>15.9%</td>
</tr>
<tr>
<td>Product Management</td>
<td>15.8%</td>
</tr>
<tr>
<td>Data Science</td>
<td>11.8%</td>
</tr>
</tbody>
</table>

Source: General Assembly part-time student data (09/2016-01/2017)
*Average = the courses listed above.
Incorporating Big Data into Health Disparities Research

<table>
<thead>
<tr>
<th>Approach</th>
<th>Example</th>
<th>Exemplary Challenge</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking structured data</td>
<td>SEER-Medicare</td>
<td>Limited to ages 65 y and older</td>
<td><a href="https://healthcaredelivery.cancer.gov/seermedicare">https://healthcaredelivery.cancer.gov/seermedicare</a></td>
</tr>
<tr>
<td>Fostering citizen science</td>
<td>Our Voice project</td>
<td>Scaling up existing small projects</td>
<td><a href="http://med.stanford.edu/ourvoice.html">http://med.stanford.edu/ourvoice.html</a></td>
</tr>
<tr>
<td>Developing big longitudinal cohorts</td>
<td>NIH All of Us cohort</td>
<td>Sustaining engagement</td>
<td><a href="https://allofus.nih.gov">https://allofus.nih.gov</a></td>
</tr>
<tr>
<td>Mining Internet and social media data to improve disparities surveillance</td>
<td>Google searches for disease and disparity keywords</td>
<td>Validation, big data hubris, algorithm dynamics and stability</td>
<td><a href="https://gking.harvard.edu/files/gking/files/0314policyforumf.pdf">https://gking.harvard.edu/files/gking/files/0314policyforumf.pdf</a> <a href="https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122963">https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122963</a></td>
</tr>
</tbody>
</table>

Note. NIH = National Institutes of Health; SEER = Surveillance, Epidemiology, and End Results program.

Source: Breen N, Jackson JS, Wood F, Wong DWS, Zhang X, Translational Health Disparities Research in a Data-Rich World, AJPH Supplement 1, 2019
HDPulse combines data sources to offer quick access to state and county data

HDPulse
An Ecosystem of Minority Health and Health Disparities Resources
Explore the Portal.

https://hdpulse.nimhd.nih.gov/
Publications to Encourage Use of Big Data in Health Disparities Research

Publications:


• Breen N, Jackson JS, Wood F, Wong DWS, Zhang X, Translational Health Disparities Research in a Data-Rich World, AJPH Supplement 1, 2019

• Bagby SP, Martin D, Chung ST, Rajapakse N, From the Outside In: Biological Mechanisms Linking Social and Environmental Exposures to Chronic disease and to Health Disparities, *AJPH*, 2019

Forthcoming: Common Data Elements-
Social Determinants of Health

**Common Data Element** - standard measure for a variable
- precisely specified question
- fixed set of permissible answers (discrete or continuous)
- Core/key variables – not all variables
- Used across multiple sites, projects, initiatives, etc.

**PhenX Toolkit** – web-based catalog of high-priority measures for genome-wide studies
- CDEs for SDOH can facilitate cross-study analyses
- Validate studies or combine to increase statistical power
<table>
<thead>
<tr>
<th>Title</th>
<th>Notice Number</th>
<th>Organization</th>
<th>Application Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAIN Initiative: Theories, Models and Methods for Analysis of Complex Data from the Brain (R01 Clinical Trial Not Allowed)</td>
<td>RFA-EB-17-005</td>
<td>NIBIB</td>
<td>9/4/2019</td>
</tr>
<tr>
<td>Secondary Data Analysis to Examine Long-Term and/or Potential Cross-Over Effects of Prevention Interventions: What are the Benefits for Preventing Mental Health Disorders? (R01 Clinical Trial Not Allowed)</td>
<td>RFA-MH-20-110</td>
<td>NIMH</td>
<td>1/3/2020</td>
</tr>
<tr>
<td>Mechanism for Time-Sensitive Opportunity in Environmental Health Sciences (R21)</td>
<td>RFA-ES-16-005</td>
<td>NIEHS</td>
<td>9/4/2019</td>
</tr>
<tr>
<td>BRAIN Initiative: Integration and Analysis of BRAIN Initiative Data (R01 Clinical Trial Not Allowed)</td>
<td>RFA-MH-19-147</td>
<td>NIMH</td>
<td>3/5/2021</td>
</tr>
<tr>
<td>BRAIN Initiative: Theories, Models and Methods for Analysis of Complex Data from the Brain (R01 Clinical Trial Not Allowed)</td>
<td>RFA-EB-17-005</td>
<td>NIBIB</td>
<td>9/4/2019</td>
</tr>
<tr>
<td>Title</td>
<td>Notice Number</td>
<td>Organization</td>
<td>Application Due Date</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Disparities Elimination through Coordinated Interventions to Prevent</td>
<td>RFA-HL-20-004</td>
<td>NHLBI</td>
<td>10/2/2019</td>
</tr>
<tr>
<td>and Control Heart and Lung Disease Risk (DECIPHeR) - Research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coordinating Center (RCC) (U24 Clinical Trial Not Allowed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia (VCID) in the United States Including in Health Disparities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Populations (U19 Clinical Trial not Allowed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration, Algorithm Development and Operations Management Center</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(U24 Clinical Trial Not Allowed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural Products NMR Open Data Exchange (NP-NODE) (U24</td>
<td>RFA-AT-19-002</td>
<td>NCCIH</td>
<td>4/18/2019</td>
</tr>
<tr>
<td>Clinical Trial Not Allowed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Methodology Development (R03 Clinical Trial Not Allowed)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Notice Number</td>
<td>Organization</td>
<td>Application Due Date</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------</td>
<td>--------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>NLM Research Grants in Biomedical Informatics and Data Science (R01 Clinical Trial Optional)</td>
<td>PAR-18-896</td>
<td>NLM</td>
<td>9/8/ 2021</td>
</tr>
<tr>
<td>Data Science Research: Personal Health Libraries for Consumers and Patients (R01 Clinical Trial Optional)</td>
<td>PAR-19-072</td>
<td>NLM</td>
<td>7/31/2021</td>
</tr>
<tr>
<td>NLM Career Development Award in Biomedical Informatics and Data Science (K01)</td>
<td>PAR-16-204</td>
<td>NLM</td>
<td>5/18/2019</td>
</tr>
</tbody>
</table>
Connect With NIMHD

Visit us online www.nimhd.nih.gov

Connect with us on Facebook www.facebook.com/NIMHD

Follow us on Twitter @NIMHD

durande@mail.nih.gov