Advancing Health Equity and Fairness in Big Data Applications in Oncology

Scarlett Lin Gomez, PhD

Department of Epidemiology & Biostatistics
Helen Diller Family Comprehensive Cancer Center
Greater Bay Area Cancer Registry

10/28/2019
Health equity means social justice in health (i.e., no one is denied the possibility to be healthy for belonging to a group that has historically been economically/socially disadvantaged).”

Paula Braveman, Public Health Rep 2014
Cells-to-Society model
A framework for multilevel research
Centers for Population Health and Health Disparities, NIH

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social conditions &amp; policies</strong></td>
<td>Poverty, norms, prejudice</td>
</tr>
<tr>
<td><strong>Institutional context</strong></td>
<td>The health care system, economic system, media, political system</td>
</tr>
<tr>
<td><strong>Social context</strong></td>
<td>Collective efficacy, social capital, racial/ethnic integration</td>
</tr>
<tr>
<td><strong>Social relationships</strong></td>
<td>Social networks, social support, religious participation, civic engagement</td>
</tr>
<tr>
<td><strong>Physical context</strong></td>
<td>Building quality, pollution, sidewalks, parks</td>
</tr>
<tr>
<td><strong>Individual demographics</strong></td>
<td>Age, race/ethnicity, SES, health insurance, education, acculturation</td>
</tr>
<tr>
<td><strong>Individual risk behaviors</strong></td>
<td>Smoking, alcohol, diet, sleep</td>
</tr>
<tr>
<td><strong>Biological responses</strong></td>
<td>Obesity, depression, stress, hypertension, chronic lung disease</td>
</tr>
<tr>
<td><strong>Biologic/Genetic pathways</strong></td>
<td>Allostatic load, biologic processes, genetic ancestry, genetic mechanisms</td>
</tr>
</tbody>
</table>

Institutional context
- The health care system, economic system, media, political system

Social context
- Collective efficacy, social capital, racial/ethnic integration

Social relationships
- Social networks, social support, religious participation, civic engagement

Physical context
- Building quality, pollution, sidewalks, parks

Individual demographics
- Age, race/ethnicity, SES, health insurance, education, acculturation

Individual risk behaviors
- Smoking, alcohol, diet, sleep

Biological responses
- Obesity, depression, stress, hypertension, chronic lung disease

Biologic/Genetic pathways
- Allostatic load, biologic processes, genetic ancestry, genetic mechanisms

Warnecke et al. AJPH 2008
Cells-to-Society model

**Fundamental causes**

They are considered fundamental causes because their influence is solely reflected at the population level in the variation in rates of disease or poor health... Their roots are embedded in policy, shared social norms about health and social practices, socioeconomic disadvantage, and policies that affect public availability of health services, including who receives them and the level and quality of service. They are the determinants of inequities rather than differences.

Warnecke et al. AJPH 2008
Cells-to-Society model

Social Determinants

Social determinant of health are the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life.

- World Health Organization

| **Social conditions & policies** | Poverty, norms, prejudice |
| **Institutional context** | The health care system, economic system, media, political system |
| **Social context** | Collective efficacy, social capital, racial/ethnic integration |
| **Social relationships** | Social networks, social support, religious participation, civic engagement |
| **Physical context** | Building quality, pollution, sidewalks, parks |
| **Individual demographics** | Age, race/ethnicity, SES, health insurance, education, acculturation |
| **Individual risk behaviors** | Smoking, alcohol, diet, sleep |
| **Biological responses** | Obesity, depression, stress, hypertension, chronic lung disease |
| **Biologic/Genetic pathways** | Allostatic load, biologic processes, genetic ancestry, genetic mechanisms |

Warnecke et al. AJPH 2008
Cells-to-Society model

**Fundamental Causes + Social Determinants**

Meta-analysis of ~50 studies found that social factors accounted for \( \frac{1}{3} \) of total deaths in US

-S. Galea AJPH 2010

Warnecke et al. AJPH 2008

---

**Social conditions & policies**
- Poverty, norms, prejudice

**Institutional context**
- The health care system, economic system, media, political system

**Social context**
- Collective efficacy, social capital, racial/ethnic integration

**Social relationships**
- Social networks, social support, religious participation, civic engagement

**Physical context**
- Building quality, pollution, sidewalks, parks

**Individual demographics**
- Age, race/ethnicity, SES, health insurance, education, acculturation

**Individual risk behaviors**
- Smoking, alcohol, diet, sleep

**Biological responses**
- Obesity, depression, stress, hypertension, chronic lung disease

**Biologic/Genetic pathways**
- Allostatic load, biologic processes, genetic ancestry, genetic mechanisms

---

**Fundamental causes**
- Social & physical context

---

**Disparate health outcomes**
- Individual factors
- Biologic factors
Study Example #1: Lung cancer in Asian American female never smokers
Lung Cancer Incidence in AANHPI* females

- AANHPI diverse population
- Lung cancer incidence varies widely across AANHPI groups
- In select case series: >50% of lung cancer cases in AANHPI females are never smokers
  - In contrast to 15% of lung cancer among all U.S. females
- Until now, no sufficiently large data source to document incidence by detailed race/ethnicity and smoking status

*AANHPI: Asian American, Native Hawaiian, Pacific Islander
Objectives

Assemble multilevel integrated dataset

Conduct incidence analysis by smoking status and detailed racial/ethnic groups

Among AANHPI female never smokers:

Assess risk associated with reproductive factors, body size, previous lung diseases, infections, air pollution
# Objectives

Assemble multilevel integrated dataset

| Electronic health records | • Sutter Health Northern California  
<table>
<thead>
<tr>
<th></th>
<th>• Kaiser Permanente Hawaii</th>
</tr>
</thead>
</table>
| Cancer registry           | • California Cancer Registry  
|                          | • Hawaii Tumor Registry |
| Neighborhood environment  | • California Neighborhoods Data System, derived for Hawaii  
|                          | • Census block group and census tract |
| Environmental exposures   | • Regional air pollution  
|                          | • Traffic density |
## Cohort description – Female lung cancer cases (N=3867)

<table>
<thead>
<tr>
<th>Group</th>
<th>Cases (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AANHPI</td>
<td>613</td>
</tr>
<tr>
<td>NHPI</td>
<td>201</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>160</td>
</tr>
<tr>
<td>Other Pacific Islander</td>
<td>41</td>
</tr>
<tr>
<td>Asian</td>
<td>412</td>
</tr>
<tr>
<td>Chinese</td>
<td>75</td>
</tr>
<tr>
<td>Filipina</td>
<td>80</td>
</tr>
<tr>
<td>Japanese</td>
<td>74</td>
</tr>
<tr>
<td>Other Asian (single group)</td>
<td>67</td>
</tr>
<tr>
<td>Multiple Asian</td>
<td>116</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>1489</td>
</tr>
<tr>
<td>Black</td>
<td>91</td>
</tr>
<tr>
<td>Hispanic</td>
<td>81</td>
</tr>
</tbody>
</table>

### Smoking status

- **Never (LCINS)**
- **Ever**
- **Missing**

<table>
<thead>
<tr>
<th>Group</th>
<th>Cases (N)</th>
<th>Never (LCINS)</th>
<th>Ever</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>AANHPI</td>
<td>613</td>
<td>38%</td>
<td>58%</td>
<td>14%</td>
</tr>
<tr>
<td>NHPI</td>
<td>201</td>
<td>15%</td>
<td>79%</td>
<td>14%</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>160</td>
<td>14%</td>
<td>58%</td>
<td>21%</td>
</tr>
<tr>
<td>Other Pacific Islander</td>
<td>41</td>
<td>20%</td>
<td>53%</td>
<td>24%</td>
</tr>
<tr>
<td>Asian</td>
<td>412</td>
<td>20%</td>
<td>58%</td>
<td>24%</td>
</tr>
<tr>
<td>Chinese</td>
<td>75</td>
<td>21%</td>
<td>58%</td>
<td>24%</td>
</tr>
<tr>
<td>Filipina</td>
<td>80</td>
<td>24%</td>
<td>58%</td>
<td>21%</td>
</tr>
<tr>
<td>Japanese</td>
<td>74</td>
<td>24%</td>
<td>58%</td>
<td>21%</td>
</tr>
<tr>
<td>Other Asian (single group)</td>
<td>67</td>
<td>24%</td>
<td>58%</td>
<td>21%</td>
</tr>
<tr>
<td>Multiple Asian</td>
<td>116</td>
<td>21%</td>
<td>58%</td>
<td>24%</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>1489</td>
<td>14%</td>
<td>38%</td>
<td>48%</td>
</tr>
<tr>
<td>Black</td>
<td>91</td>
<td>14%</td>
<td>38%</td>
<td>48%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>81</td>
<td>14%</td>
<td>38%</td>
<td>48%</td>
</tr>
</tbody>
</table>
Cohort description – Female lung cancer cases

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Cases (N)</th>
<th>Never (LCINS)</th>
<th>Ever</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>AANHPI</td>
<td>613</td>
<td>15%</td>
<td>38%</td>
<td>48%</td>
</tr>
<tr>
<td>NHPI</td>
<td>201</td>
<td>14%</td>
<td>20%</td>
<td>66%</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>160</td>
<td>20%</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>Other Pacific Islander</td>
<td>41</td>
<td>14%</td>
<td>20%</td>
<td>66%</td>
</tr>
<tr>
<td>Asian</td>
<td>412</td>
<td>79%</td>
<td>53%</td>
<td>18%</td>
</tr>
<tr>
<td>Chinese</td>
<td>75</td>
<td>40%</td>
<td>53%</td>
<td>27%</td>
</tr>
<tr>
<td>Filipina</td>
<td>80</td>
<td>24%</td>
<td>53%</td>
<td>23%</td>
</tr>
<tr>
<td>Japanese</td>
<td>74</td>
<td>79%</td>
<td>53%</td>
<td>18%</td>
</tr>
<tr>
<td>Other Asian (single group)</td>
<td>67</td>
<td>50%</td>
<td>53%</td>
<td>27%</td>
</tr>
<tr>
<td>Multiple Asian</td>
<td>116</td>
<td>21%</td>
<td>40%</td>
<td>39%</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>1489</td>
<td>14%</td>
<td>38%</td>
<td>48%</td>
</tr>
<tr>
<td>Black</td>
<td>91</td>
<td>20%</td>
<td>38%</td>
<td>42%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>81</td>
<td>14%</td>
<td>38%</td>
<td>48%</td>
</tr>
</tbody>
</table>
Incidence of **never-smoking lung cancer among females**

<table>
<thead>
<tr>
<th></th>
<th>AAIR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AANHPI</td>
<td>17.1</td>
<td>(14.9, 19.9)</td>
</tr>
<tr>
<td>NHPI</td>
<td>15.2</td>
<td>(10.2, 22.5)</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>16.7</td>
<td>(10.5, 26.4)</td>
</tr>
<tr>
<td>Other Pacific Islander</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>17.5</td>
<td>(15.0, 21.0)</td>
</tr>
<tr>
<td>Chinese</td>
<td>22.8</td>
<td>(17.3, 35.9)</td>
</tr>
<tr>
<td>Filipina</td>
<td>20.1</td>
<td>(14.1, 31.5)</td>
</tr>
<tr>
<td>Japanese</td>
<td>6.4</td>
<td>(3.6, 28.7)</td>
</tr>
<tr>
<td>Other Asian (single group)</td>
<td>20.3</td>
<td>(13.4, 32.1)</td>
</tr>
<tr>
<td>Multiple Asian</td>
<td>22.2</td>
<td>(16.1, 32.4)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>10.1</td>
<td>(9.0, 11.6)</td>
</tr>
<tr>
<td>Black</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.5</td>
<td>(5.7, 13.0)</td>
</tr>
</tbody>
</table>
Ongoing work

Risk analysis in six exposure domains

1. Smoking
2. Previous lung disease
3. Infectious disease
4. Reproductive factors
5. Body size
6. Environment
   - Air pollution
   - Neighborhood

EHR

Geocoded residential address
Linked to PM2.5 & neighborhood factors

Machine learning – iterative random forests (M. DeRouen)
Joint contributions of multilevel factors to LCINS risk
Study Example #2: Prostate cancer in African American men
Why prostate cancer in African American men?

<table>
<thead>
<tr>
<th>Rate/100,000</th>
<th>African American</th>
<th>Non-Hispanic White</th>
<th>Ratio AA:NHW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence</td>
<td>179.2</td>
<td>101.7</td>
<td>1.8</td>
</tr>
<tr>
<td>Mortality</td>
<td>39.8</td>
<td>18.1</td>
<td>2.2</td>
</tr>
</tbody>
</table>
RESPOND Study: Prostate cancer in African American men

Addressing the role of social stressors, genetics, and tumor genomics & biology

RESPOND is an integrated set of studies all focused on a common theme of identifying the multi-level determinants and characteristics of aggressive prostate cancer in African American men.
Recruitment and research sites

Recruitment primarily through cancer registries
Projects & Integration

Project 1
Survey and Geospatial Data

Project 2
Germline DNA

Project 3
Somatic Tumor

Project 4
TME Tumor

RESPOND Cohort

10,000 PCa cases identified through Cancer Registries in seven states

Multilevel Stressors
Project 1

Somatic Events
Project 3

Tumor Inflammation
Project 4

Germline Risk
Project 2

Low, Intermediate, High-Risk Prostate Cancer

Prostate Cancer Recurrence and Survival

Prospective Follow-up of RESPOND (beyond year 5)
Project 1 Data & Analysis Flowchart

All AA & WH PCa cases (dx 2000-2013)
N=149,000 AA
N=668,000 WH

Cancer registry data

Geospatial data (blk gp-, tract-level)

Neighborhood stressors

Aim 1
Stressors → Aggressive disease
Stressors → Mortality
N=149,000 AA, 668,000 WH PCa cases
Project 1 Data & Analysis Flowchart

All AA & WH PCa cases (dx 2000-2013)  
N=149,000 AA  
N=668,000 WH

Cancer registry data  
Neighborhood stressors

Geospatial data (blk gp-, tract-level)

Aim 1  
Stressors → Aggressive disease  
Stressors → Mortality  
N=149,000 AA, 668,000 WH PCa cases

RESPOND cohort (dx 2015-2018)  
N=10,000 AA PCa cases

Geospatial data (latitude/longitude-, blk grp-, tract-level)  
Public-record data

Survey data + cancer registry data  
Life course individual and neighborhood stressors

Aim 2  
Life course individual, neighborhood stressors → Aggressive disease  
N=10,000 AA PCa cases
Project 1 Data & Analysis Flowchart

Aim 1
Stressors → Aggressive disease
Stressors → Mortality
N=149,000 AA, 668,000 WH PCa cases

Aim 2
Life course individual, neighborhood stressors → Aggressive disease
N=10,000 AA PCa cases

Aim 3a
Genetic ancestry → Stressors
N=10,000 AA PCa cases

Aim 3b
Stressors → Somatic profiles
N=3000 AA PCa cases

Aim 3c
Stressors + somatic profiles + germline genetics → Aggressive disease
N=3000 AA PCa cases

All AA & WH PCa cases (dx 2000-2013)
N=149,000 AA
N=668,000 WH

Cancer registry data

Geospatial data (blk gp-, tract-level)

Neighborhood stressors

RESPOND cohort (dx 2015-2018)
N=10,000 AA PCa cases

Survey data + cancer registry data

Geospatial data (latitude/longitude-, blk grp-, tract-level)

Public-record data

Lifecourse individual and neighborhood stressors

Project 2: Germline genetics

Project 3: Somatic profiles

Project 1 Data & Analysis Flowchart

African American Prostate Cancer Study

Responding to the Cancer

UCSF
Survey topics

**Sociodemographic background**: multiple racial/ethnic groups, birthplace, education, health insurance, financial security, marital status, sexual identity, income, employment

**Lifecourse multi-level social stressors**: neighborhood factors, lifetime and everyday discrimination, medical mistrust, adverse childhood experiences, “John Henryism” active coping

**Lifestyle**: body size, physical activity, alcohol use, smoking history
Structural racism – tract level only for MSAs (K. Beyer)

- Red-lining index
- Index of racial bias in mortgage lending

Racial/ethnic segregation

- City- or MSA-level indices of relative distributions between Blacks and Whites – e.g., dissimilarity index
- Local segregation measures (J. Gibbons) - typology of various racial/ethnic combinations in census tract
- Location quotient of relative concentration of AAs in a block group compared to the MSA (K. Beyer)

Racial/ethnic composition

- % Black in census tract
Big Data, SDOH, Health Equity: A Cautionary Tale
Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Heidi Ledford
“It is vital to develop tools that move from assessing individual risk to evaluating the production of risk by institutions so that, ultimately the public can hold them accountable for harmful outcomes.”

Recommendations
Big Data + Multilevel SDOH → Health Equity

Health inequities arise from fundamental causes & multilevel SDOH
- Study structural & institutional factors
- Collect address data & integrate neighborhood context
- Lifecourse, cumulative effects, weathering

Data needs
- Standardized measures
- Granular data
- Training for healthcare professionals, toolkits

Include small populations
- Consortia comprising multilevel data
- Integrative data analysis

Pay attention to intersectionality
- Interactions among multiple marginalized SDOH statuses
Dreamlab.ucsf.edu
Disparities Research Environment And oMics

@ slingomez
Scarlett.gomez@ucsf.edu