UCSF
University of California
San Francisco

# 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

@slingomez
#NatlCancerForum

**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)."
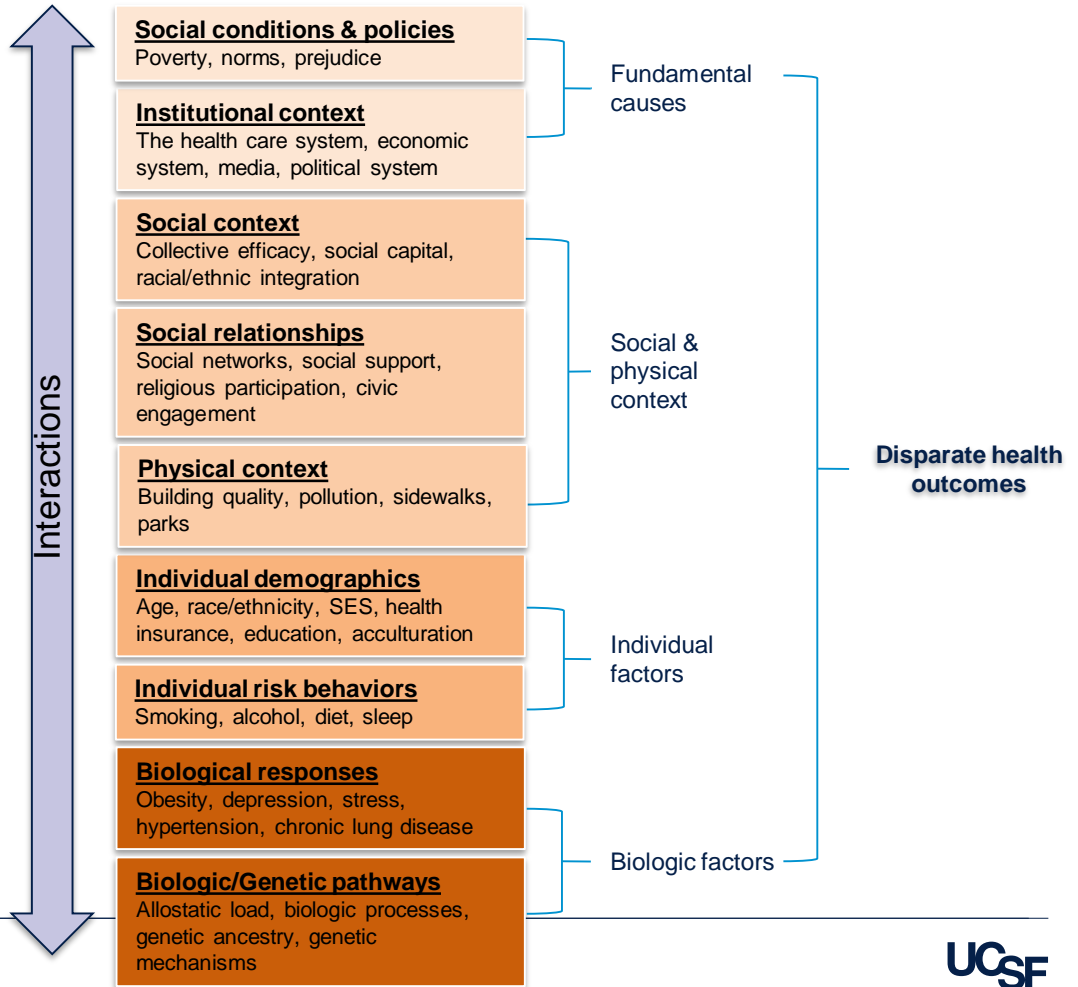
**Paula Braveman, Public Health Rep 2014**

**DREA.ı.ı.**LAB

UCSF

Cells-to-Society model

A framework for multilevel research

Centers for Population Health and Health Disparities, NIH

Warnecke et al. AJPH 2008

**Interactions**

**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

Individual factors

Biologic factors

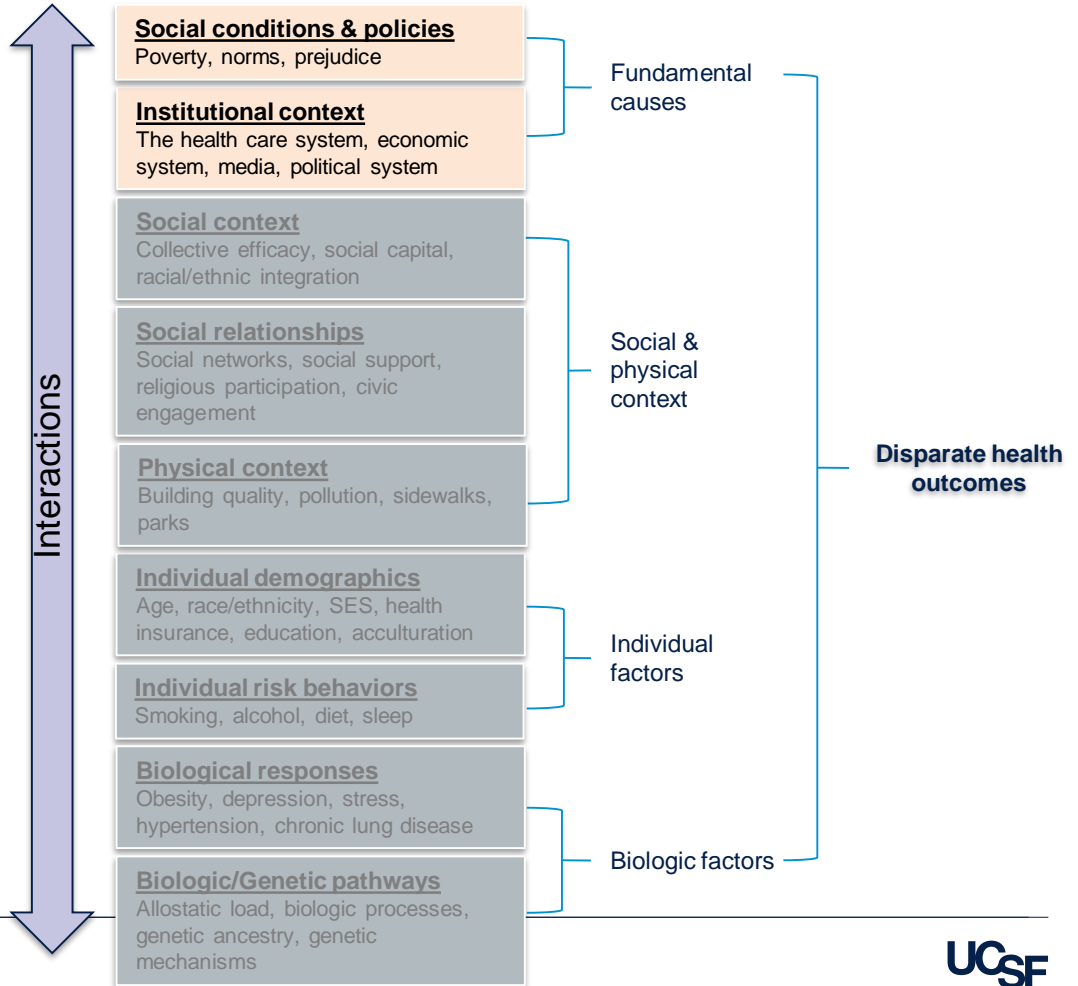**Disparate health outcomes**

# 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

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

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

Fundamental causes

**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

Social & physical context

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

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

Individual factors

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

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

Biologic factors

Interactions

**Disparate health outcomes**
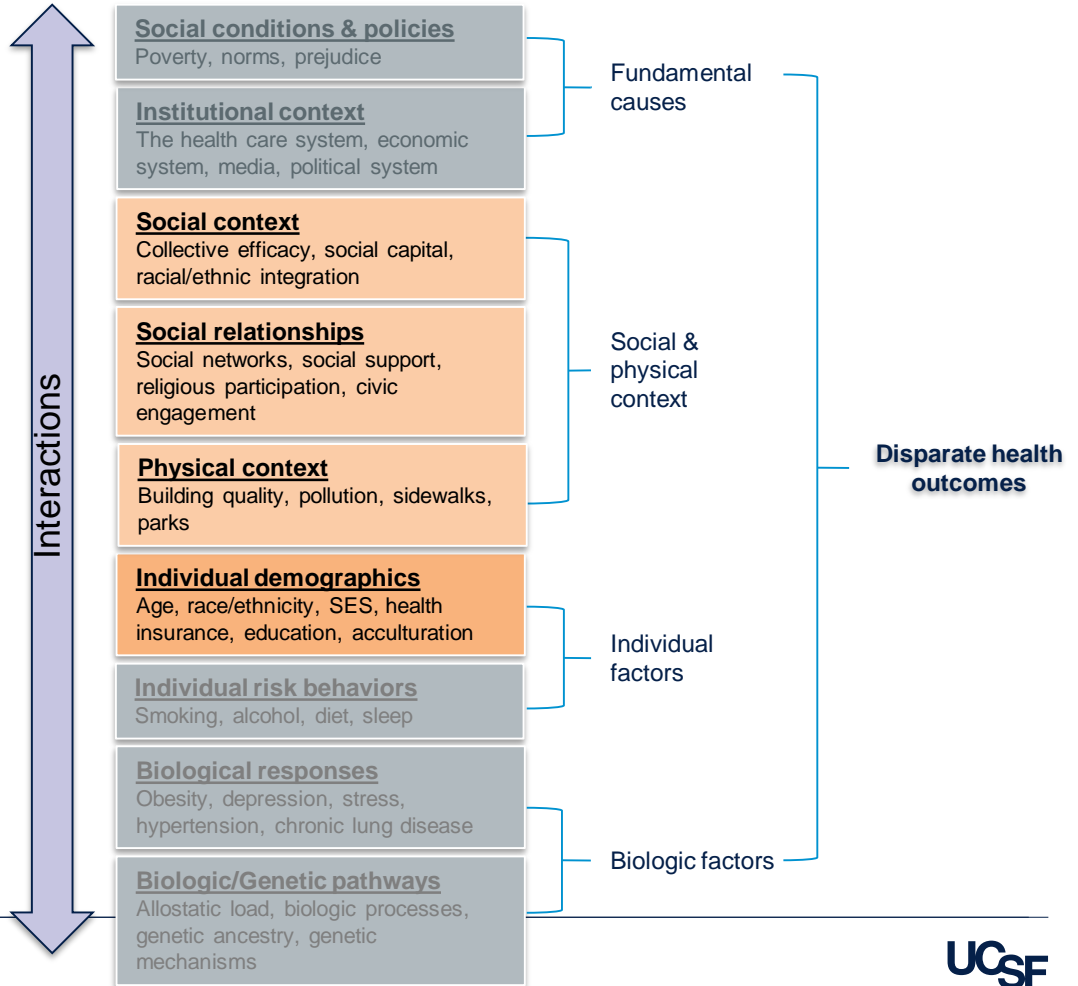
DREA.l.l.LAB

UCSF

# 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.
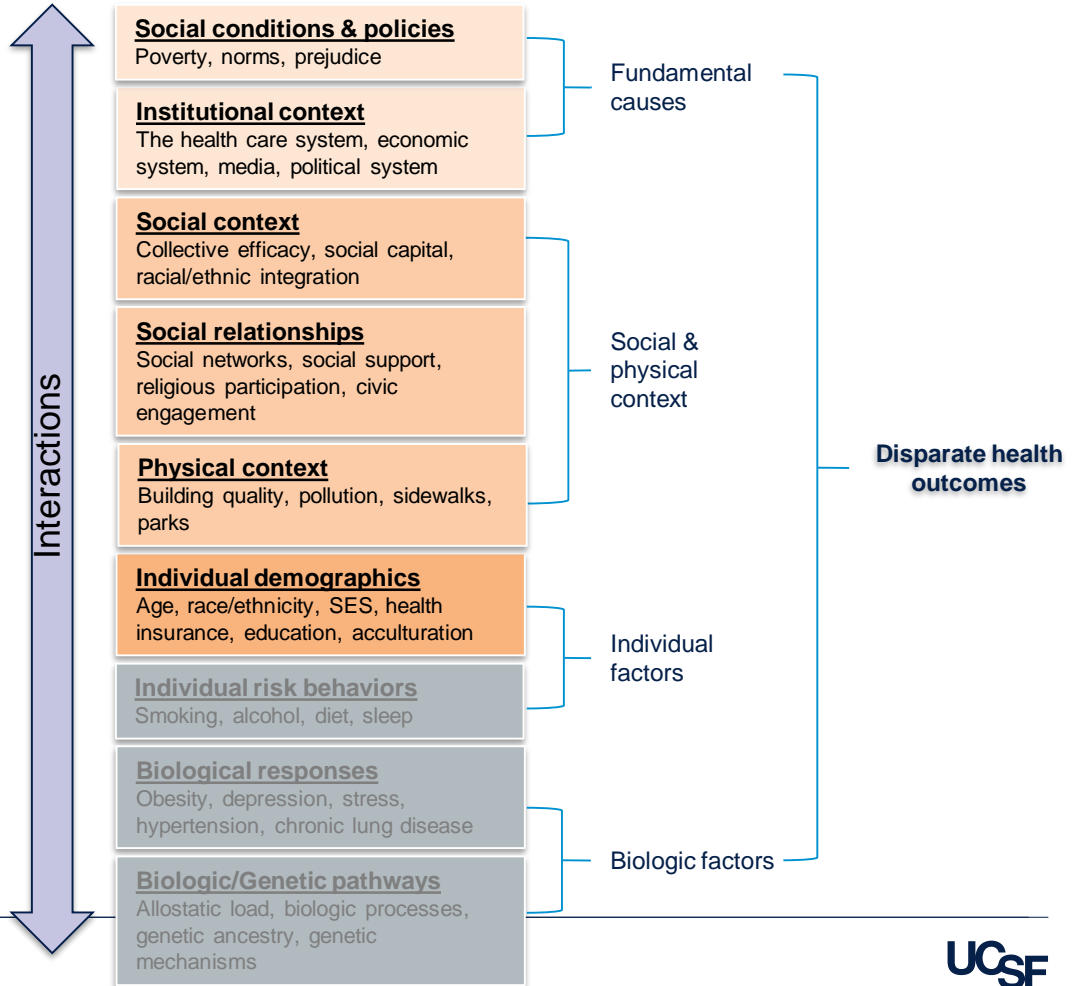-World Health Organization

Warnecke et al. AJPH 2008

**Interactions**

**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

Individual factors

Biologic factors

**Disparate health outcomes**

**DREA.l.l.LAB**

UCSF

# Cells-to-Society model

## Fundamental Causes + Social Determinants

Meta-analysis of ~50 studies found that social factors accounted for **1/3 of total deaths** in US

-S. Galea AJPH 2010

Warnecke et al. AJPH 2008

**Interactions**

**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

Individual factors

Biologic factors

**Disparate health outcomes**

**DREA.l.l.LAB**

UCSF

# Study Example #1: Lung cancer in Asian American female never smokers

DREA.lıl.LAB

UCSF

# 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

**DREA.ılı.**LAB

UCSF

# 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

**DREA.I.I.**LAB

UC SF

# Objectives

## Assemble multilevel integrated dataset

| Electronic health records | • Sutter Health Northern California<br>• Kaiser Permanente Hawaii |
|---|---|
| Cancer registry | • California Cancer Registry<br>• Hawaii Tumor Registry |
| Neighborhood environment | • California Neighborhoods Data System, derived for Hawaii<br>• Census block group and census tract |
| Environmental exposures | • Regional air pollution<br>• Traffic density |

DREA.li.LAB

UCSF

# Cohort description – Female lung cancer cases (N=3867)

| | Cases (N) |
|---|---|
| AANHPI | 613 |
| NHPI | 201 |
| Native Hawaiian | 160 |
| Other Pacific Islander | 41 |
| Asian | 412 |
| Chinese | 75 |
| Filipina | 80 |
| Japanese | 74 |
| Other Asian (single group) | 67 |
| Multiple Asian | 116 |
| Non-Hispanic White | 1489 |
| Black | 91 |
| Hispanic | 81 |

Smoking status

■ Never (LCINS) ■ Ever ■ Missing

AANHPI: 38%
NHPI: 15%
Native Hawaiian: 14%
Other Pacific Islander: 20%
Asian: 50%
Chinese: 79%
Filipina: 53%
Japanese: 24%
Other Asian (single group): 58%
Multiple Asian: 40%
Non-Hispanic White: 21%
Black: 14%
Hispanic: 38%

0%  20%  40%  60%  80%  100%

DREA.l.l.LAB

# Cohort description – Female lung cancer cases



| | Cases (N) |
|---|---|
| AANHPI | 613 |
| NHPI | 201 |
| Native Hawaiian | 160 |
| Other Pacific Islander | 41 |
| Asian | 412 |
| Chinese | 75 |
| Filipina | 80 |
| Japanese | 74 |
| Other Asian (single group) | 67 |
| Multiple Asian | 116 |
| Non-Hispanic White | 1489 |
| Black | 91 |
| Hispanic | 81 |

Smoking status
- Never (LCINS)
- Ever
- Missing

| Group | Never (LCINS) |
|---|---|
| AANHPI | 38% |
| NHPI | 15% |
| Native Hawaiian | 14% |
| Other Pacific Islander | 20% |
| Asian | 50% |
| Chinese | 79% |
| Filipina | 53% |
| Japanese | 24% |
| Other Asian (single group) | 58% |
| Multiple Asian | 40% |
| Non-Hispanic White | 21% |
| Black | 14% |
| Hispanic | 38% |

DREA.ılı.LAB

# Incidence of never-smoking lung cancer among females

**Age-adjusted incidence rate (per 100,000)**

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



**DREA.l.l.LAB**
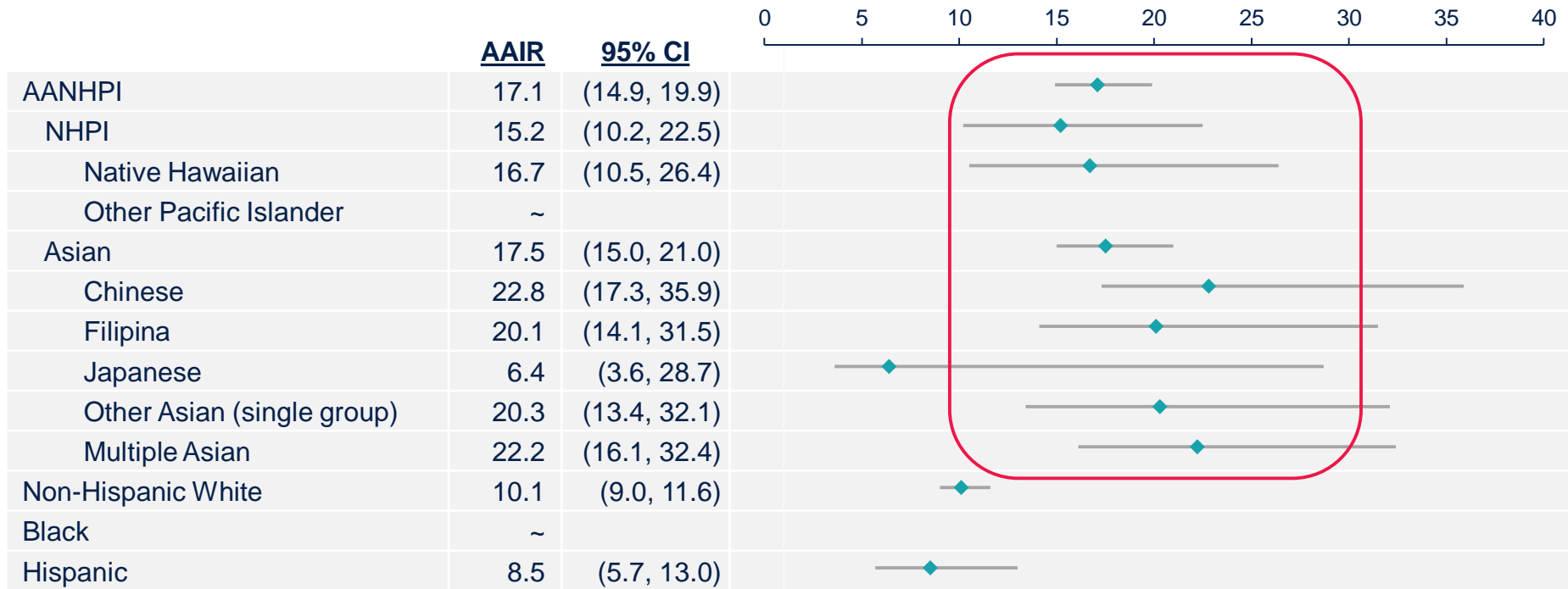
# Ongoing work

Risk analysis in six exposure domains

1. Smoking
2. Previous lung disease
3. Infectious disease — EHR
4. Reproductive factors
5. Body size
6. Environment
    Air pollution — Geocoded residential address
    Neighborhood — Linked to PM2.5 & neighborhood factors

Machine learning – iterative random forests (M. DeRouen)
    Joint contributions of multilevel factors to LCINS risk

DREA.ı.ı.LAB                                                    UCSF

# Study Example #2: Prostate cancer in African American men

**DREA.l.l.LAB**

UCSF

# Why prostate cancer in African American men?

| Rate/100,000 | African American | Non-Hispanic White | Ratio AA:NHW |
|---|---|---|---|
| Incidence | 179.2 | 101.7 | 1.8 |
| Mortality | 39.8 | 18.1 | 2.2 |

DREA.l.l.LAB

UCSF

# 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.*

**RESPOND**
African American Prostate Cancer Study

U19CA214253 (Haiman)

**DREA.ı.ı.LAB**

UCSF

# Recruitment and research sites
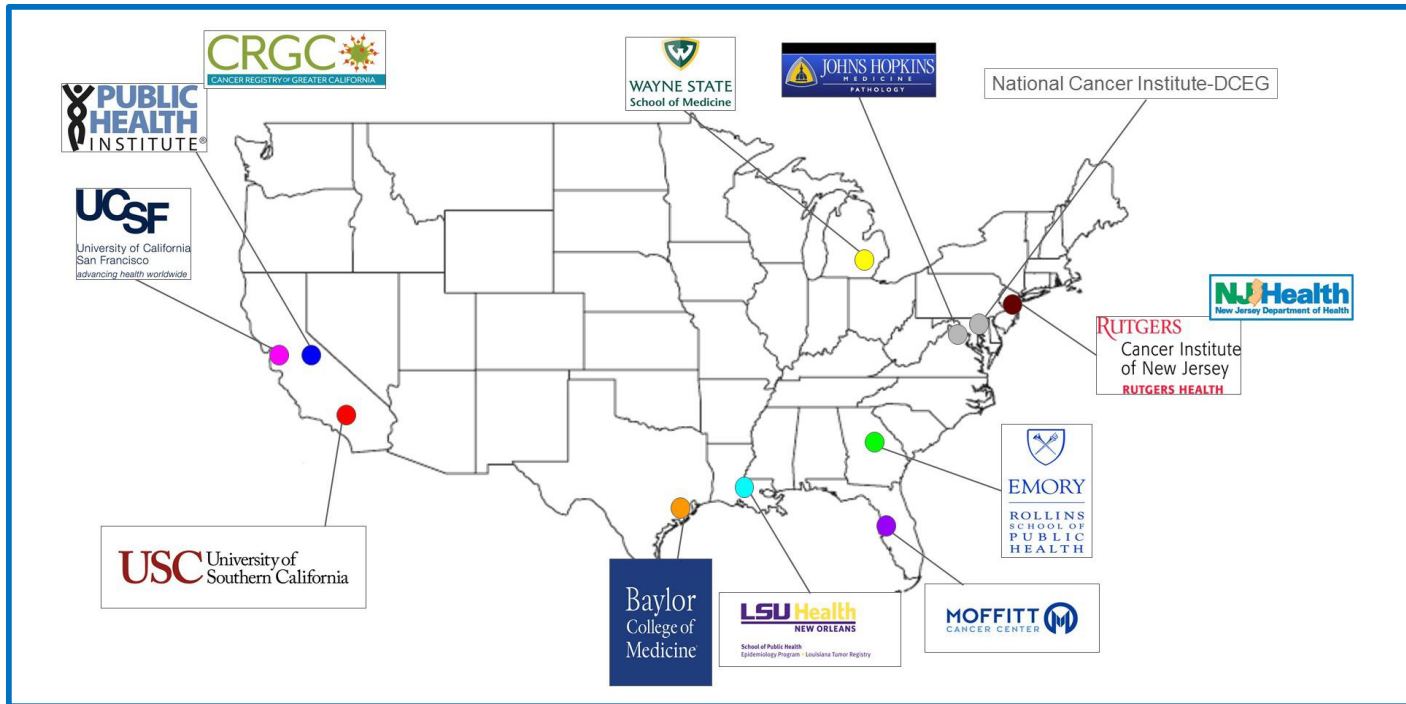


Recruitment primarily through cancer registries

Project 1
Survey and Geospatial Data
n=10,000

Project 2
Germline DNA
n=7,500

Project 3
Somatic Tumor
n=3,000

Project 4
TME Tumor
n=1,000

RESPOND Cohort

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

Multilevel Stressors
**Project 1**

Somatic Events
**Project 3**

Germline Risk
**Project 2**

Tumor Inflammation
**Project 4**

Low, Intermediate, High-Risk Prostate Cancer

Prostate Cancer Recurrence and Survival

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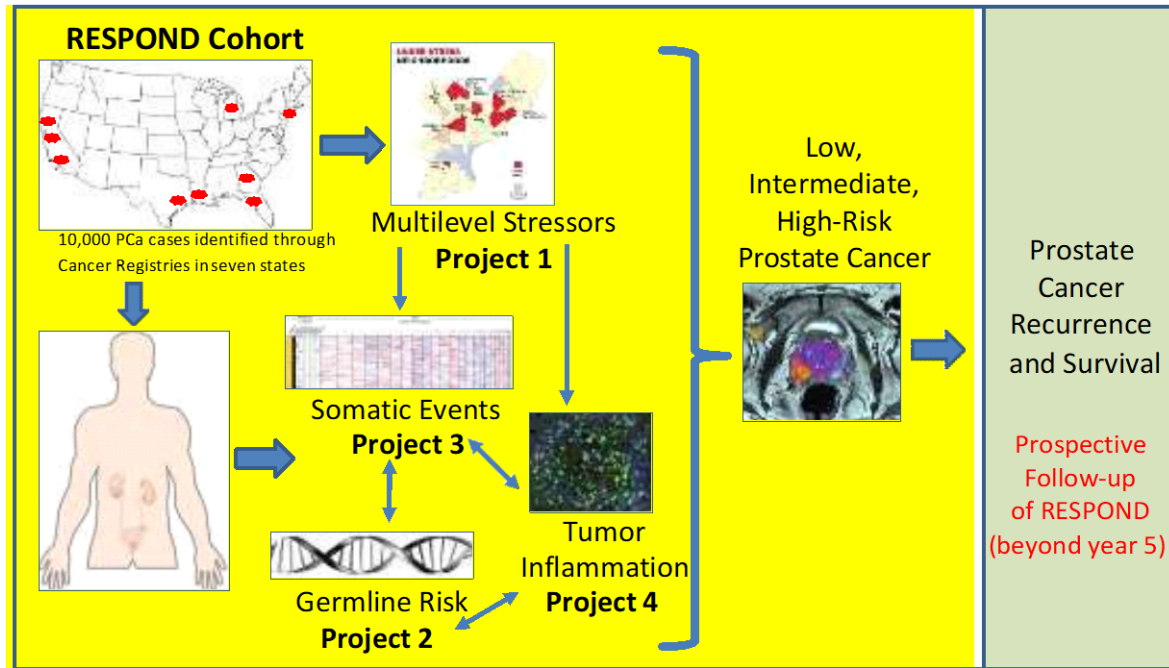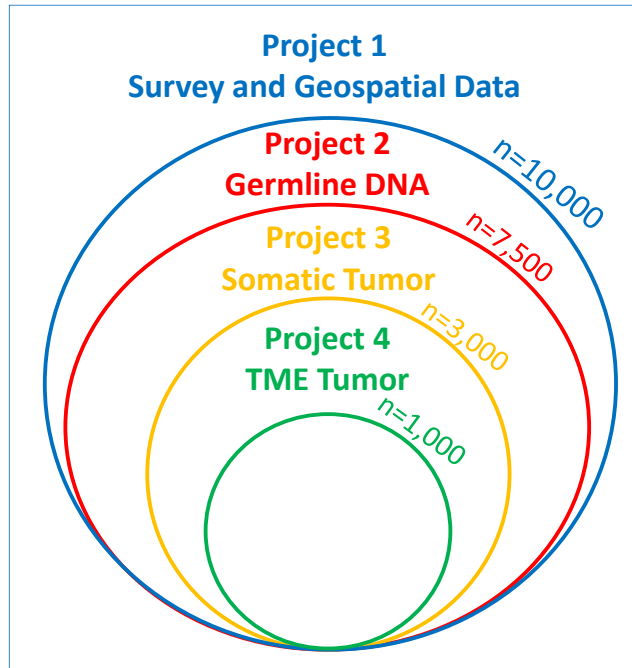
Projects & Integration
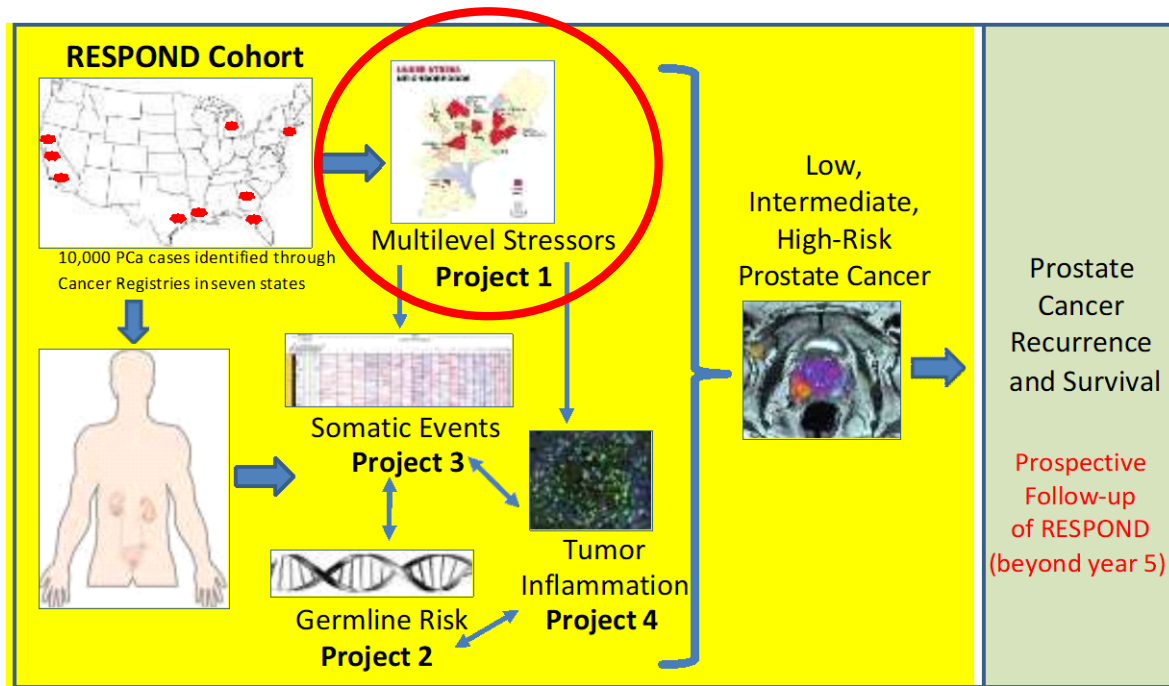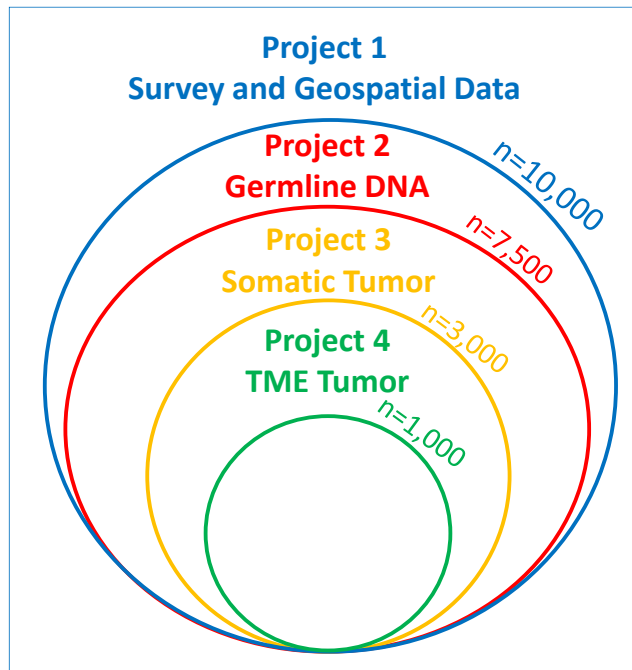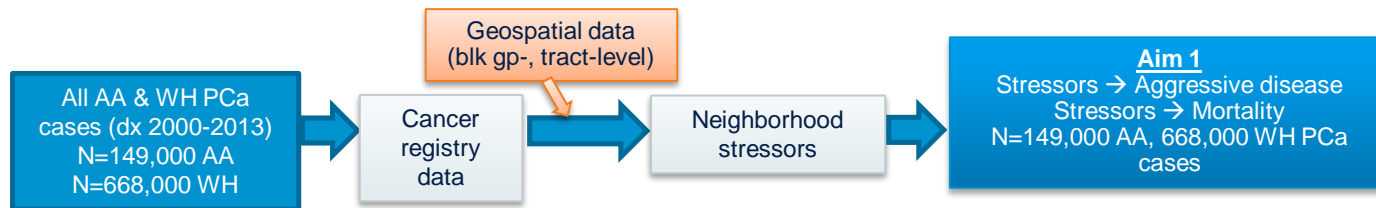
Project 1 Data & Analysis Flowchart
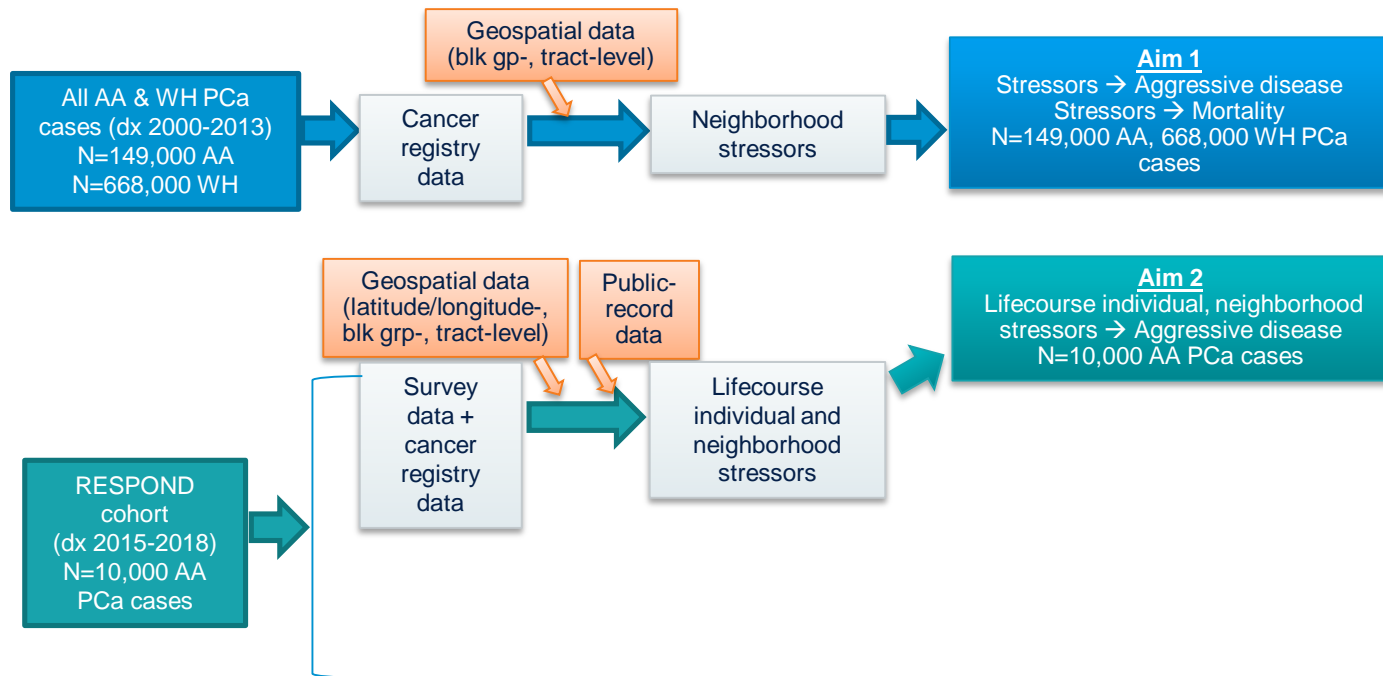
# Project 1 Data & Analysis Flowchart

RESPOND
African American Prostate Cancer Study

**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

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

Public-record data

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

Survey data + cancer registry data

→ Lifecourse individual and neighborhood stressors

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

DREA.l.l.LAB

UCSF

Project 1 Data & Analysis Flowchart

# 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

**DREA.ili.LAB**

UCSF

# Big Data, SDOH, Health Equity: A Cautionary Tale

**DREA.l.l.LAB**

UCSF

MENU

# nature

Subscribe

Search

Login

# 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**

29

DREA.l.i.LAB

UCSF

MENU

# nature

Subscribe

Search

Login

**NEWS** · 24 OCTOBER 2019

**Millions**
**health-c**

Study reveals ra
highlights ways

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."

-Ruha Benjamin, "Assessing risk, automating racism", Science, Oct 25, 2019

DREA.lıl.LAB

UCSF

# Recommendations

**DREA.ıiı.**LAB

# 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

DREA.I.I.LAB                                                                    UCSF

Dreamlab.ucsf.edu
Disparities Research Environment And oMics

DREA.li.LAB

UCSF

@slingomez
Scarlett.gomez@ucsf.edu