Developing a Diverse Research Community
SESSION 3:

Health Disparities Research and Electronic Health Records: Considerations and Methods

Charisse Madlock-Brown, PhD

About Me

Research interests: observational research, HIT interventions, multimorbidity, health disparities, social determinants of health, and COVID-19

Defining Health Disparities

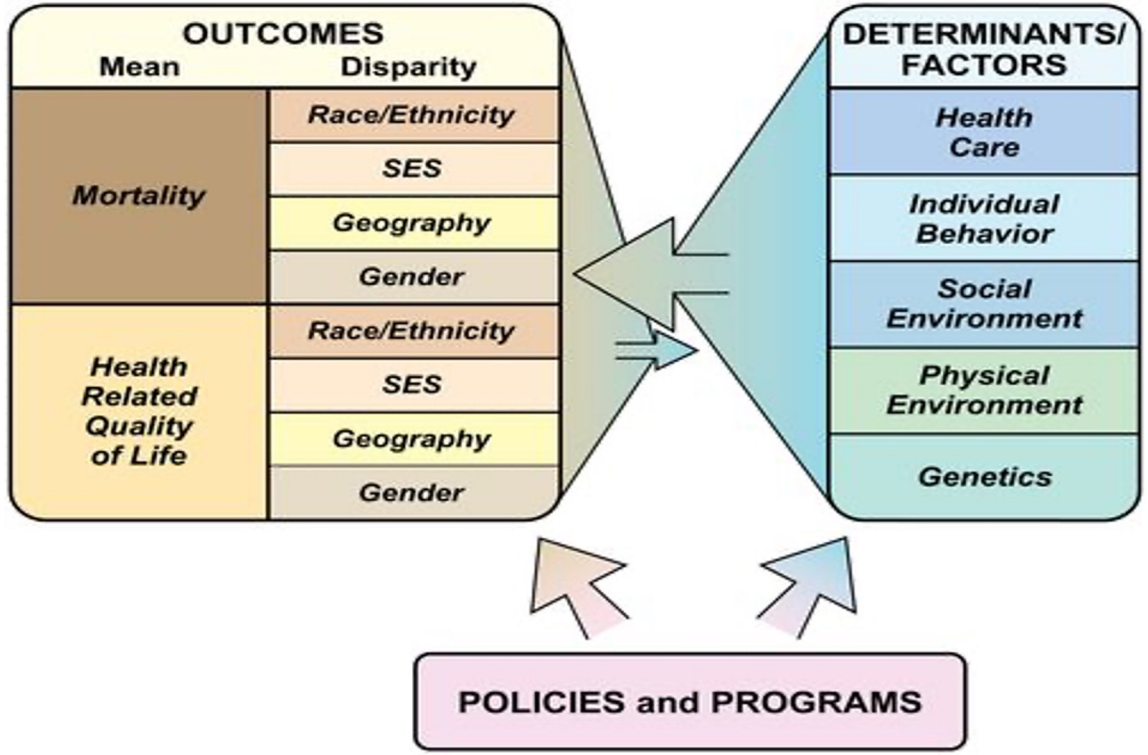
"The difference in health outcomes for defined disadvantages populations that are worse than the White reference population"

-- National Institute on Minority Health and Health Disparities

Populations with health disparities according to NIH

- Racial and ethnic minority groups
- People with lower socioeconomic status (SES)
- Underserved rural communities
- Sexual and gender minority (SGM) groups

Health Disparities In Context



From, “What are Population Health Determinants or Factors?”
-www.improvingpopulationhealth.org

Health Disparities Research using EHRs

- First step is understanding how to do this research in a principled way
  - Collaboration with researchers from target population and communities
  - Understanding social drivers for disparities
  - Understanding how documentation and utilization patterns bias research

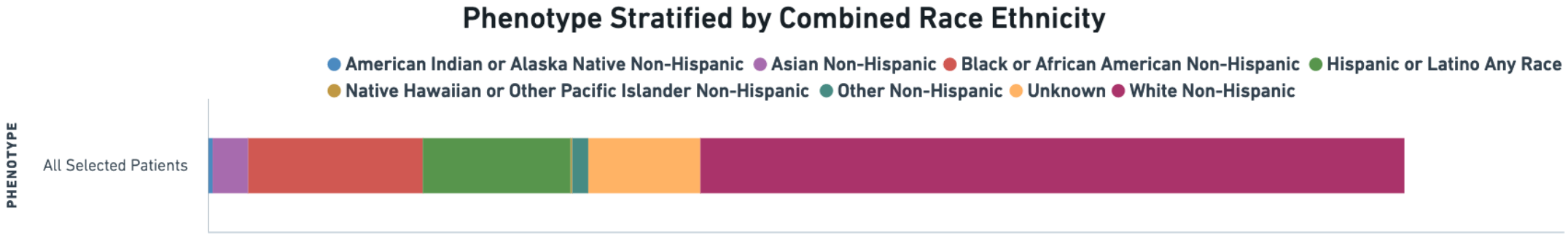
Health Disparities Research using EHRs

Understanding how to get NIH funding

- Current directions
- Tips from the research “grape-vine”

Research Data Warehouse Example

- National COVID Cohort Collaborative



More data, More Problems

HIGH VARIETY DATA, AND MEASUREMENT ERROR

HIGH VOLUME DATA, AND ANALYTIC RIGOR

HIGH VELOCITY DATA, AND INTERVENTION OPTIMIZATION

Where the issue starts: racial categorization

Defining Race

- “Any one of the groups that human beings are often divided into based on **physical traits or ancestry.** Race **is a culturally and politically charged term,** for which definitions and meaning are context-specific. It is related to individual and/or group identity, and is often **linked to stereotypes of visible physical attributes such as skin and hair pigmentation.”**

*Social Determinants of Health Factors for Gene–Environment COVID-19 Research: Challenges and Opportunities. *Advanced Genetics*

Racial Categories Data issues

Table 1. Office of Management and Budget (OMB) revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, 1997.

OMB category	HL7[a] code	Category definition
Race: American Indian or Alaska Native	1002-5	A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment
Race: Asian	2028-9	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam
Race: Black or African American	2054-5	A person having origins in any of the black racial groups of Africa. Terms such as "Haitian"; or "Negro"; can be used in addition to "Black or African American"
Race: Native Hawaiian or Other Pacific Islander	2076-8	A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands
Race: White	2106-3	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa
Ethnicity: Hispanic or Latino	2135-2	A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race. Ethnicity is considered a distinct category from race

*Issues With Variability in Electronic Health Record Data About Race and Ethnicity: Descriptive Analysis of the National COVID Cohort Collaborative Data Enclave. *JMIR Medical Informatics.*

Racial Categories Vary by Institutions

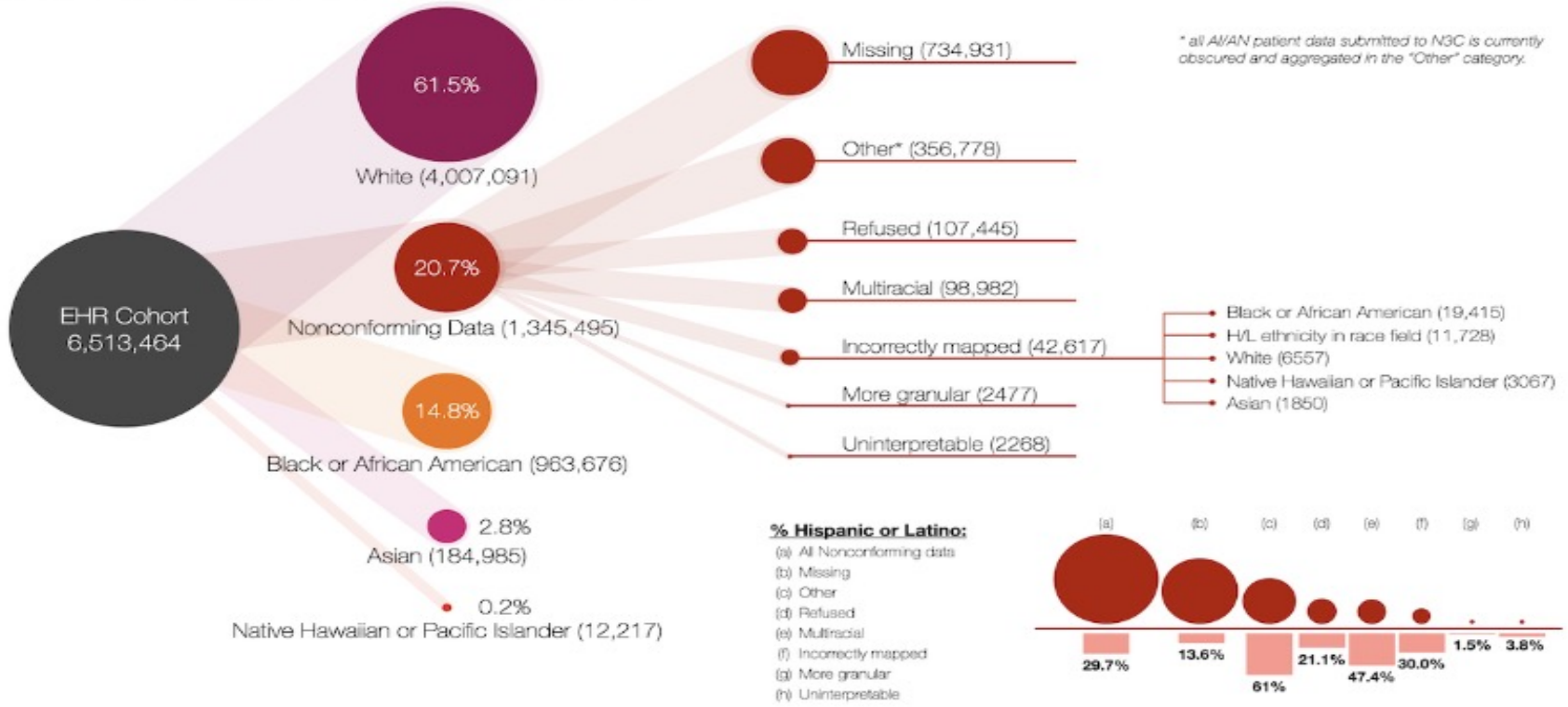
Figure 1. Race data reporting schema by contributing sites. Although data partners did contribute data on American Indian or Alaska Native patients, as noted elsewhere, these data were intentionally obscured. OMB: Office of Management and Budget.

Reporting schema	Data partners, *n* (%)	American Indian or Alaska Native*	Asian	Black or African American	Native Hawaiian or Other Pacific Islander	White	Black	Hispanic	Asian Indian	Chinese	Filipino	Japanese	Korean	Other Pacific Islander	Polynesian	Vietnamese	
A	41 (73.2)		✓	✓	✓	✓											Standard only
B	2 (3.6)		✓	✓	✓	✓								✓			
C	1 (1.8)		✓	✓	✓	✓		✓									Standard with additional categories
D	1 (1.8)		✓	✓	✓	✓			✓								
E	1 (1.8)		✓	✓	✓	✓			✓	✓	✓	✓	✓		✓	✓	
F	5 (8.9)		✓	✓		✓											
G	2 (3.6)			✓		✓											Missing standard
H	1 (1.8)			✓	✓	✓											
I	1 (1.8)		✓			✓	✓										Missing standard, additional categories
J	1 (1.8)		✓	✓		✓								✓			
Total data partners, *n*		0	53	55	49	56	1	1	2	1	1	1	1	3	1	1	
		Standard OMB Categories					Other race categories										

*Issues With Variability in Electronic Health Record Data About Race and Ethnicity: Descriptive Analysis of the National COVID Cohort Collaborative Data Enclave. *JMIR Medical Informatics.*

Conformance issues and Racial Categories

Figure 3. Weighted tree diagram of nonconforming race data. AI/AN: American Indian or Alaska Native; EHR: electronic health record; H/L: Hispanic/Latino; N3C: National COVID Cohort Collaborative.



*Issues With Variability in Electronic Health Record Data About Race and Ethnicity: Descriptive Analysis of the National COVID Cohort Collaborative Data Enclave. *JMIR Medical Informatics.*

Long COVID, Selection Criteria, and Representation

Long COVID: Phenotyping and Estimating Prevalence

- Varied nature of symptomology
- Data collection methods lead to different symptom representation (e.g., self reporting vs. Problem lists)
- Prevalent estimates from 15-50 percent

*How common is long COVID? Why studies give different answers. *Nature*

Challenges with EHR Systems

- EHR systems have limitations such as potential disparities in data availability and biases related to selection criteria.
- Sicker patients are more likely to seek care.
- Healthcare-seeking behavior varies by gender, race/ethnicity, and socio-economic status

*Hidden in plain sight: Bias towards sick patients when sampling patients with sufficient electronic health record data for research. BMC Medical Informatics and Decision Making.

*Race, Medical Mistrust, and Segregation in Primary Care as Usual Source of Care: Findings from the Exploring Health Disparities in Integrated Communities Study. Journal of Urban Health.

Selection Criteria AND Long COVID Studies

- Selection criteria
  - Medical history and COVID-19-related hospitalization
- these criteria may also impact the representativeness of the data
  - As data completeness can affect a patient's likelihood of meeting these requirements. For example, limiting to patients with a recorded COVID-19 diagnosis may bias against women with long COVID.

Study Goals for Work in Progress

- The goal of this study is to assess the impact of inclusion criteria
  - on representation along gender, race, healthcare-seeking, and socio-economic lines.
- the study aims to shed light on the limitations of EHR systems and how they impact the generalizability of long COVID studies.
- Additionally, by modeling the impact of selection criteria, solutions can be assessed

Study Cohort and Variables

- cohort selection: Patients with a U09.9 diagnoses code
- Independent variables: COVID-19 dx, COVID-19 related hospitalization (get the definition from logic liaison team), and history (difference in days between first observation and initial long-covid diagnosis).
- Outcome variables: is_female, is_white.

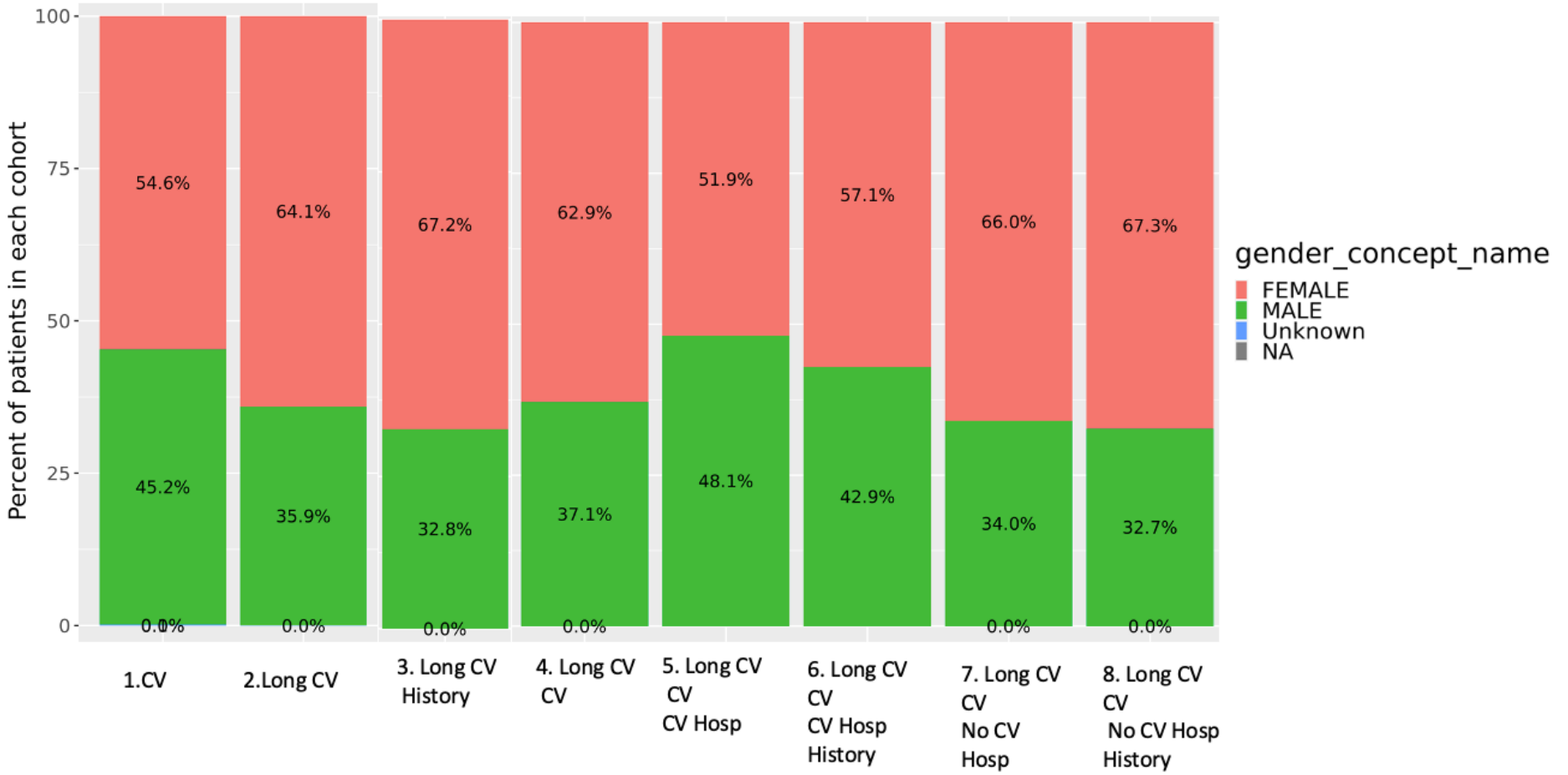
Modeling Selection Criteria Impact on Representation

- Mixed effect Logistic regression for binary variables.
- Example: is_female~covid_dx + covid_hospitalization + months_of_data + (1|data_partner_id)

Gender Breakdown across Race

	American Indian or Alaska Native Non-Hispanic	Asian Non-Hispanic	Black or African American Non-Hispanic	Hispanic or Latino Any Race	Unknown Non Hispanic	White Non-Hispanic	p-value
N	234	543	3581	2118	1732	19450	
female	157 (67.1)	346 (63.7)	2586 (72.2)	1368 (64.5) +-5	1094 (63.0)+-15	12317 (63.3)+-5	<0.001
male	77 (32.9)	197 (36.3)	995 (27.8)	750 (35.4)	638 (36.8)	7133 (36.7)	
Gender Unknown	<20	<20	<20	<20	<20	<20	

Gender Distribution across Cohorts



CV stands for COVID
Hosp stands for Hospital

Modeling the Impact of Selection Criteria on gender

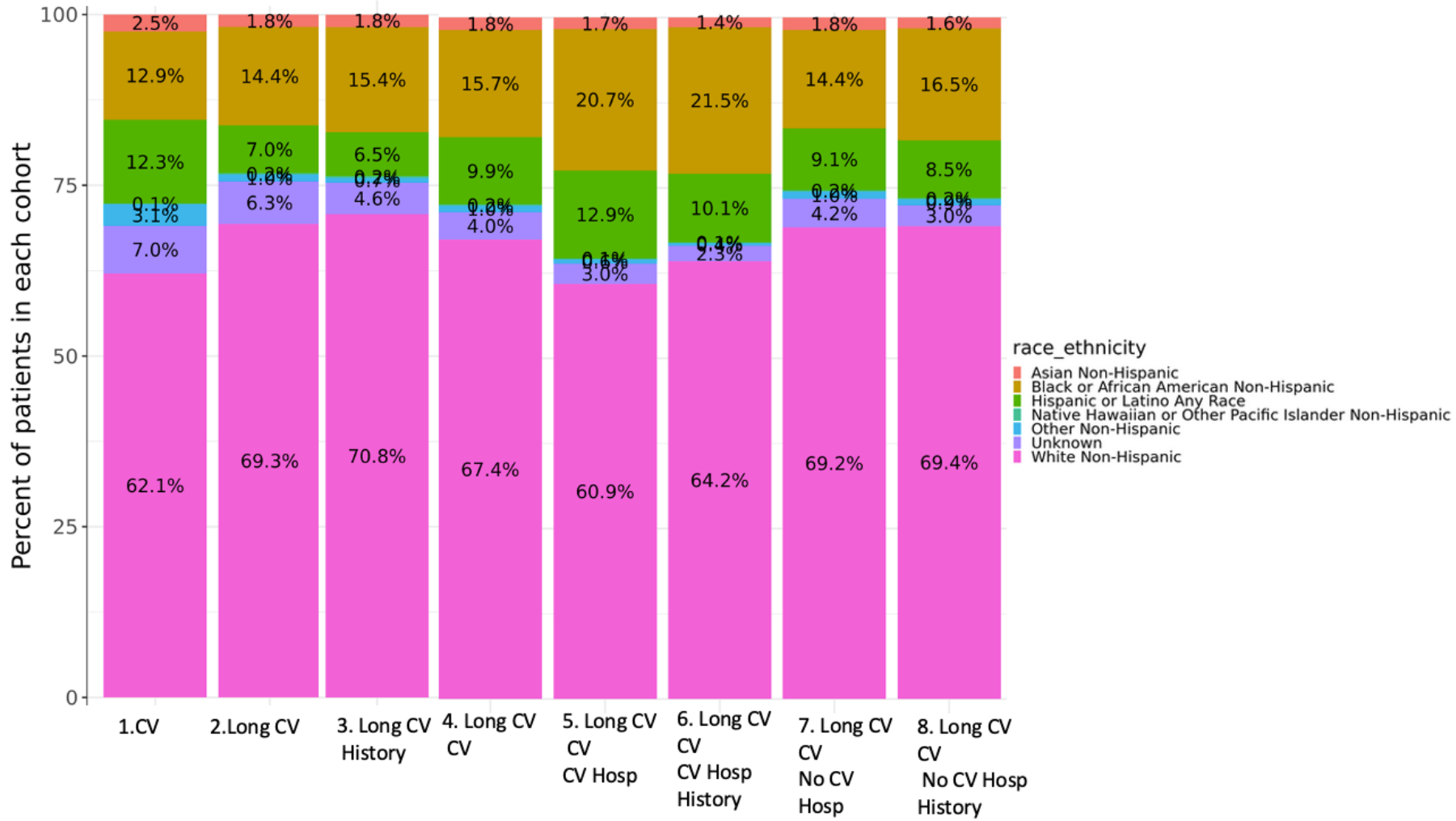
- Mixed Effect GLM: is_female~covid_dx + covid_hospitalization + months_of_data + (1|data_partner_id)
- All significant

Odds ratios with confidence interval			
Response Variable	COVID Dx	COVID-related Hospitalization	History (Months)
female	0.91 (0.86, 0.95)	0.57 (0.52, 0.62)	2.27 (2.10, 2.46)

Race/Ethnicity Breakdown

	American Indian or Alaska Native Non-Hispanic	Asian Non-Hispanic	Black or African American Non-Hispanic	Hispanic or Latino Any Race	Unknown Non Hispanic	White Non-Hispanic	p-value
N	234	543	3581	2118	1732	19450	
history 2+years	163	386	2662	1390	882	13889	<0.001
history less	71 (30.3)	157 (28.9)	919 (25.7)	728 (34.4)	850 (49.1)	5561 (28.6)	
has COVID dx	155 (66.2)	331 (61.0)	2470 (69.0)	1466 (69.2)	906 (52.3)	12178 (62.6)	<0.001
no COVID dx	79	212	1111	652	826	7272	
has COVID hospitalization	<20	50 (9.2)	539 (15.1)	259 (12.2)	119 (6.9)	1838 (9.4)	<0.001
no COVID hospitalization	229 +-10	493	3042	1859	1613	17612	

Race/Ethnicity Distribution across Cohorts



CV stands for COVID

Hosp stands for Hospital

Modeling the Impact of Selection Criteria on Race/Ethnicity

- Mixed Effect GLM: is_whte~covid_dx + covid_hospitalization + months_of_data + (1|data_partner_id)
- All significant
- Will try a multinomial model next

Odds ratios with confidence interval			
Response Variable	COVID Dx	COVID-related Hospitalization	History (Months)
White Non-Hispanic	0.88 (0.83, 0.93)	0.78 (0.72, 0.85)	1.12 (1.11, 1.32)

Engaging with Disadvantaged Communities

NIH Health Disparities Direction

- Health Disparities research is being more broadly funded across institutions
- Emphasis on Engaging with communities (e.g., AIM-Ahead initiative).
- Health disparities cannot just be tacked on to other studies
- Need to clearly show impact and novelty