Race, Ethnicity and Ancestry in Genomic Research
--- Use and Misuse ---

Tesfaye B. Mersha, PhD

Associate Professor

Cincinnati Children's Hospital Medical Center
Department of Pediatrics
University of Cincinnati College of Medicine

1

Disclosures

- **Nothing to disclose**

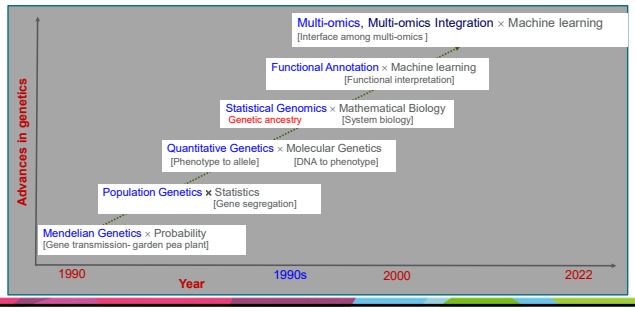
2

OUTLINE

- Population descriptors: race, ethnicity, ancestry.
- Misuse of population descriptors.
- Best practice for use of population descriptors.
- Case study: Causal inference approach
- Discussions and recommendations.

3

The Concept of Genetics is Evolving



4

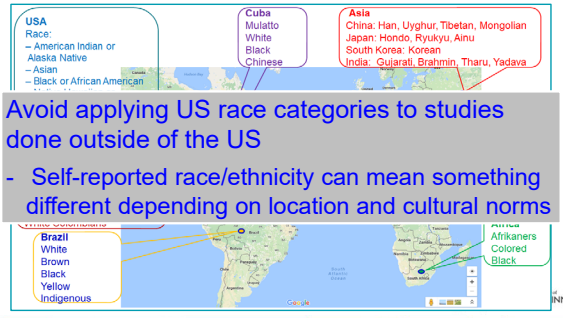
History of Human Racial Classification

Millions of Americans change their racial or ethnic identity from one census to the next BUT the concept of race and ethnicity remain the same.

➢ *How individual perceive themselves as well as how they are perceived by others.*

5

Global View of race and ethnicity: How do we harmonize?



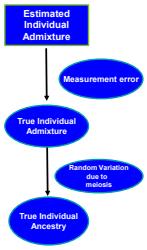
Avoid applying US race categories to studies done outside of the US

- Self-reported race/ethnicity can mean something different depending on location and cultural norms

6

Ancestry is an inferred proportion based on admixture estimate

- Ancestry is a latent or unknown variable.
  - Without complex pedigrees, the # founders is unknown
- Admixture is an error containing *proxy* measurement of ancestry.
  - $Admixture_i = Ancestry_i + Error_i$
- Error
  - Measurement – incomplete coverage of genome, missing data, etc..
  - Reference population – Imprecise founding populations
  - Biological – meiosis, etc..
  - Random – all the rest



7

Real life complexity of race, ethnicity, and ancestry

An example…

"**Mr. X** was born in Japan to Japanese parents, but as an infant, he was adopted by an Italian family in Italy. Ethnically, he feels Italian: he eats Italian food, he speaks Italian, he knows Italian history and culture. He knows nothing about Japanese history and culture. But when he comes to the United States, he is treated racially as Asian."

Japanese = Ancestry

Italian = Ethnicity

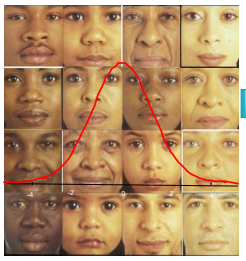
Asian = Race

Modified from Stella Yi, PhD, MPH

8

There is human diversity NOT human races

- Human populations *have not been isolated long* enough to form true biological races.
- There is no scientific way to determine race.
- Racial categories change over the years.



Forms a *gradient* rather than a racial cluster

https://www.youtube.com/watch?v=CRcex9NEJZE

9

Under what conditions should we use race?

- To identify or refer disparities related to:
  - Unequal treatment due to actual/perceived race.
  - Investigate healthcare access (diagnosis, treatment)
  - Race-specific socio-environmental risk factors.
- Patient recruitment for genetics studies to address biomedical research gaps/questions.

10

Misuse of race information in disease genetics

- Race comparison in evaluating the genetic causes of diseases without accounting for socio-environmental risk factor disparities.
  - Conveys racial disparities are caused by genetic variants rather than structural racism.
- Using race/ethnicity as a covariate to explain genetic variations (instead of social inequalities).
- Rationalizing health disparities using racial and genetic essentialism.
  - Justify racial disparities for reason of genomic studies.

11

Misuse of race category in clinical algorithms

- Automatic adjustments for race in clinical decision-making and treatment. However:
  - Race itself is a self-reported variable that cannot be accurately captured or measured.
- Example: kidney, lung, heart, pain management.

Clinical utility	Use of race
Cardiology	[redacted] to the risk score if the patient is identified as nonblack
Nephrology	Correction by a factor of [redacted] if the patient is identified as black
Urology	[redacted] are added for nonblack race
Endocrinology	[redacted] are added for nonblack race.
Pulmonology	correction factors for persons labeled as black ([redacted])
Oncology	Rank the race/ethnicity categories in the following [redacted]: white, American Indian, black, Hispanic, Asian.

(N engl j med 383: 9, 2020)

12

Faulty interpretation of preterm birth data: Genetics for preterm birth racial disparities?

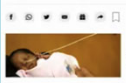
Racial disparity in the frequency of recurrence of preterm birth.

Z. Kistka, L. Palomar, +5 authors L. Muglia •
Published 2007 • Medicine •
American journal of obstetrics and gynecology

OBJECTIVE
We examined the hypothesis that black race independent of other factors increases the risk for extreme preterm birth and its frequency of recurrence at a similar gestational age.

The New York Times

Study Points to Genetics in Disparities in Preterm Births



YOUNG AND FRAGILE A girl who was born at 27 weeks being fed through a tube.

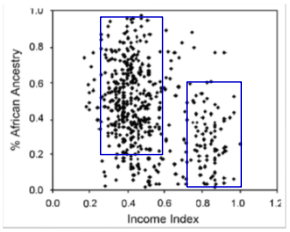
By Nicholas Bakalar
Feb. 27, 2007

Black women have significantly higher rates of premature birth than white women, and a new study suggests there may be underlying genetic factors even when other known risks are taken into account.

Am J Obstet Gynecol. 2007 Feb;196(2):131.e1-6.
https://www.nytimes.com/2007/02/27/health/27birt.html

13

Faulty interpretation of genetic ancestry: Ancestry for income gap?



Could African ancestry be a surrogate for:

- Less education
- Fewer years of parental schooling
- No inheritance
- Less wealth

European Journal of Human Genetics (2008) 16, 762–765

14

African American race is associated with AD, independent of African genetic ancestry

Odds of AD among full cohort (N=70,984)	Odds Ratio	(95% CI)	Odds of AD among the full cohort, including both genetic ancestry and self-identified race (N=70,984)	Odds Ratio	(95% CI)
			Self-identified race (African American vs white)	2.23	(1.04,4.95)
Percent African Ancestry (by 1-unit increase)	**1.01**	**(1.01, 1.01)**	**Percent African Ancestry (by 1-unit increase)**	1.00	(0.99, 1.01)
Sex (female vs male)	**1.25**	**(1.12, 1.39)**	Sex (female vs male)	**1.25**	**(1.12, 1.38)**
Income			Income		
<$20,000	ref		<$20,000	ref	
$20,000-60,000	0.85	(0.68, 1.07)	$20,000-60,000	0.85	(0.68, 1.07)
>$60,000	**0.76**	**(0.61, 0.95)**	>$60,000	0.76	(0.61, 0.95)
Education			Education		
Less than high school	ref		Less than high school	ref	
High School or equivalent	0.95	(0.66, 1.38)	High School or equivalent	0.95	(0.65, 1.37)
College or graduate school	1.05	(0.73, 1.53)	College or graduate school	1.05	(0.72, 1.52)
History of atopy	**2.19**	**(1.99, 2.42)**	History of atopy	**2.19**	**(1.99, 2.42)**
Age (each additional year)	**1.00**	**(1.00, 1.00)**	Age (each additional year)	1.00	(1.00, 1.00)

J Allergy Clin Immunol. 2020 145 (1): 192-198

15

Adjustment for socio-environmental risk factors is critical

- SES risk factors can confound the effect of ancestry on disease risk.
  - Example: **T2D** in samples from Mexico and Colombia.

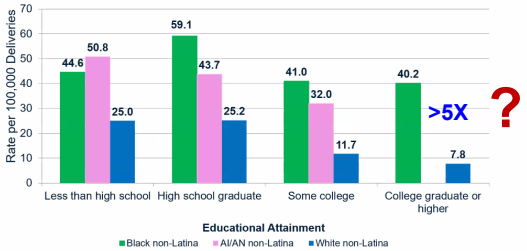
Population	Beta - Eur	p value	Beta - SES	p value	Beta - Eur (SES as cov)	p value
Mexico	-2.84	2×10^{-5}	-0.95	2×10^{-10}	- 1.79	0.02
Colombia	-1.36	0.02	-0.41	8×10^{-7}	-0.45	0.46

- **Colombian ancestry**: accounting for SES eliminated the association of European ancestry with lower risk of diabetes
- **Mexican ancestry**: accounting for SES attenuated the association.

Florez et al. 2009. Diabetologia 52:1528-1536

16

Pregnancy-related mortality ratios by educational attainment: 2007-2016 ?



Source: Petersen E et al. Racial/Ethnic Disparities in Pregnancy-Related Deaths – United States, 2007-2016. MMWR. Sept. 6, 2019. vol 68. no 35

17

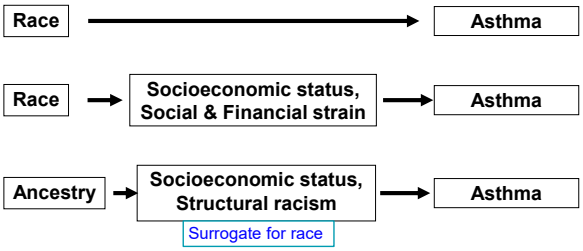
Case study

Previous studies predominately focused on:

- Direct association of self-reported race with disease outcome.
- Direct biologic effects of genetic ancestry on disease outcome.
- Here, we implement a comprehensive analysis of racial disparities in diseases using causal inference approach.

18

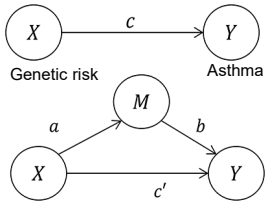
Conceptual model



19

Mediation model

- Mediation model add a putative mediator, M, to an established association model, X → Y.



Effect of *X* on *Y* without considering mediation.

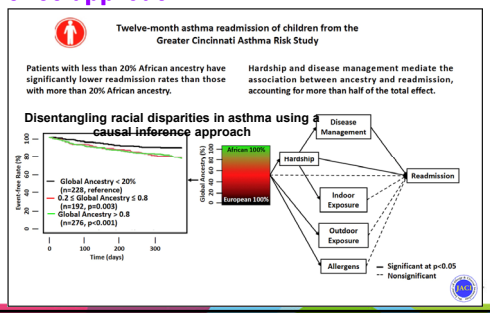
Effect of *X* on *Y* including mediation.

Figures from https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01984/full

20

Disentangling racial disparities in asthma using a causal inference approach

Mersha TB et al. J Allergy Clin Immunol. 2021. PMID: 34217757



21

Concluding remarks

- Race should not be used as proxy for biological risk factors.
- Race is a proxy for structural and institutional racism-- shapes the experiences of all socially classified groups.
- Emphasizing group differences-based on ancestry could lead to further disparities in health and healthcare outcomes.
- Racial gaps are compounded by ***inequities*** in genomic research.
- No international consensus on the definitions of race/ethnicity.
- The need to employ standard measures of race/ethnicity so you can make meaningful comparisons with worldwide datasets.

22

Recommendations

- Develop world-wide acceptable population descriptor *ontology*, like gene ontology, phenotype ontology, etc. for large scale worldwide consortia.
  - Include scientists and local lore from around the world and listen to their perception of race, ethnicity, and ancestry.
  - Develop a common language and common practices supported by empirical evidence.
- Develop worldwide population descriptors consortia including socio-environmental and cultural determinants.
  - Harmonize how race/ethnicity information is collected and defined worldwide.
- Reconsider the use of race correction in clinical algorithms.

23

Acknowledgements

Mersha Lab

- Yadu Gautam, PhD, Research Associate
- Eskezeia Dessie, PhD, Postdoctoral Fellow
- Esteban Correa, PhD, Postdoctoral Fellow
- Meenasri Kumbaji, MS, Research Assistant
- Sudhir Ghandikota, MS, Graduate student
- Siqi Chen, BS, MS Student
- Prithvi Jami, BS, MS student

Collaborators

- Neeru Hershey, MD, PhD
- Marc Rothenberg, MD, PhD
- Jonathan Bernstein, MD
- Rob Kahn, MD, MPH
- Tilahun Abebe, PhD
- Mingon Wang, PhD
- Lili Ding, PhD
- Michael Kobor, PhD
- Changchun Xie, PhD
- Andrew Beck, MD, MPH

Cincinnati Children's
National Institutes of Health
NIH
NIH
NIH
CpG

24