Environmental and Social Risk Factors in Sickle Cell Disease and NASEM Report on Use of Population Descriptors in Genomics Research

Charmaine DM Royal, PhD

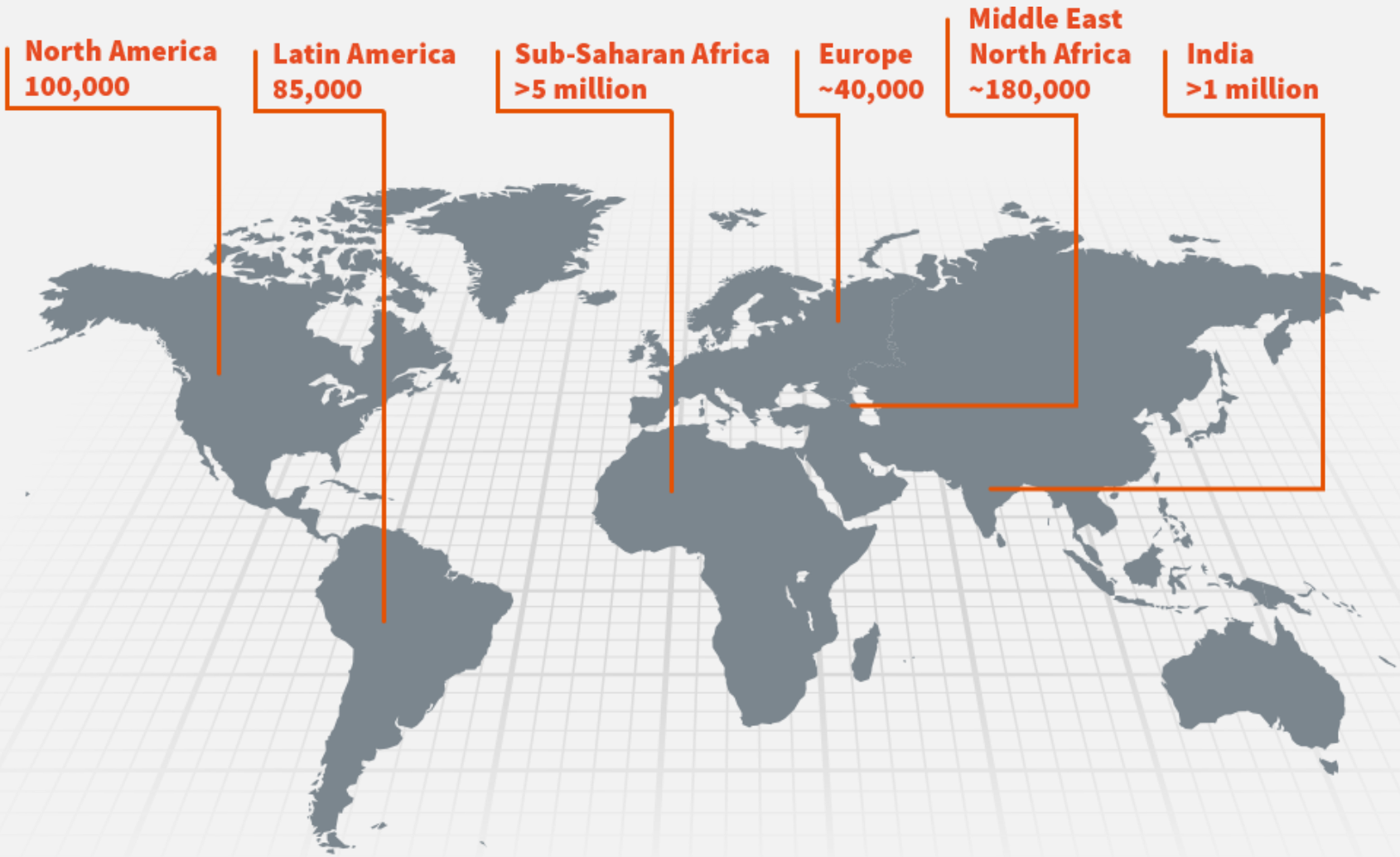
Robert O. Keohane Professor
African & African American Studies, Biology, Global Health,
and Family Medicine & Community Health

Duke

Duke GRID
GENOMICS RACE IDENTITY DIFFERENCE

Duke | CENTER *for* TRUTH, RACIAL HEALING & TRANSFORMATION

Global impact of sickle cell disease



https://www.notaloneinsicklecell.com

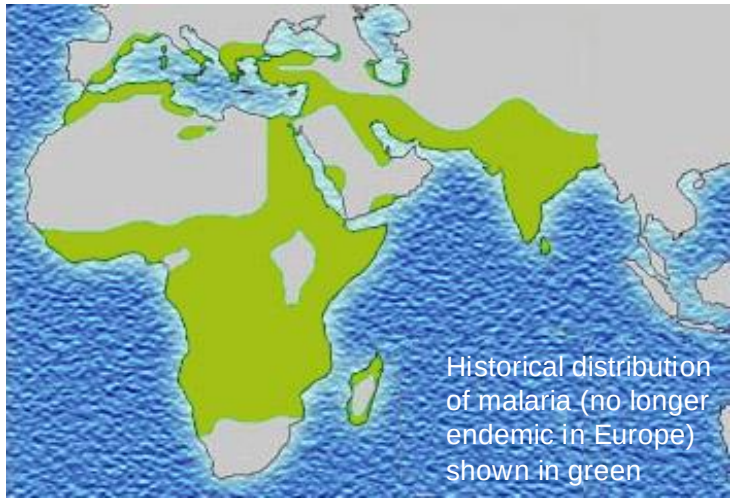
Global distribution of sickle cell disease

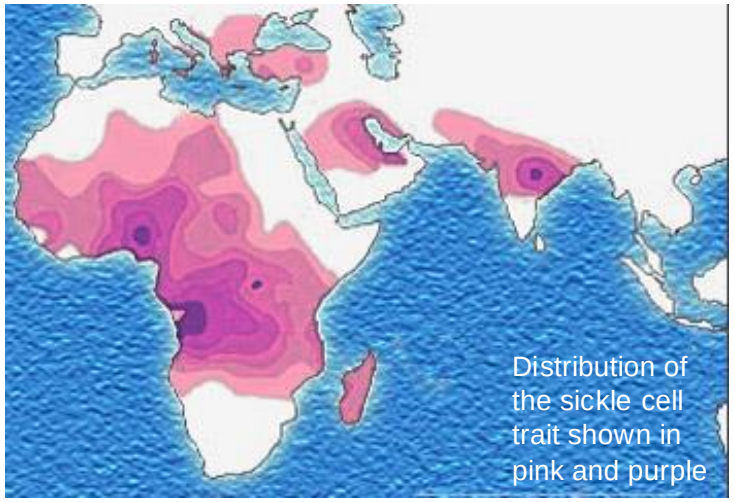
- Parts of Africa, India, Middle East, Mediterranean, Caribbean, Americas
- About 20 million people affected with SCD globally
- Worldwide more than 300,000 babies are born with SCD each year; at least 75% in Africa
- Cameroon: ~ 1:60 births
- Jamaica: ~ 1:150 births
- US: ~ 1:365 live births for black Americans and about 1:16,300 live births for Hispanics

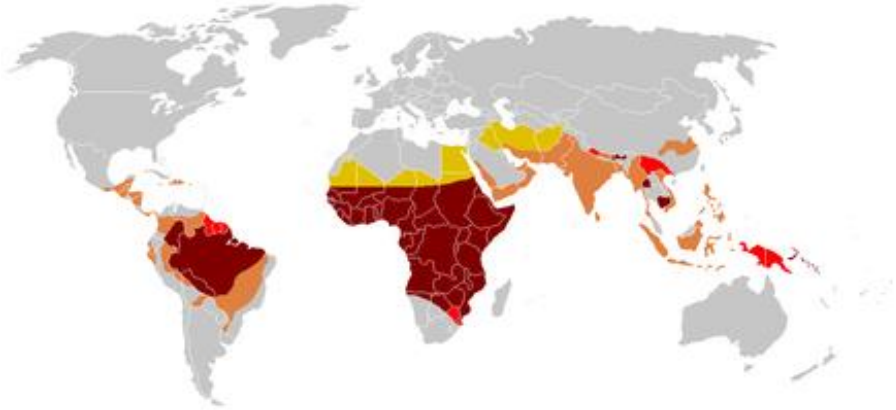
Molecular basis of sickle cell disease

- Hemoglobin S (HbS) - ~7300 years ago
- Single nucleotide substitution (G**A**G → G**T**G) in beta-globin gene on chromosome 11
- Group of genetic blood disorders characterized by sickle-shaped red blood cells
- Sickled red blood cells
  - sticky
  - rigid
  - reduced life-span

	HbA ("normal")	HbS (sickle)
DNA sequence	CCT GAG GAG GGA CTC CTC	CCT G**T**G GAG GGA C**A**C CTC
RNA sequence	CCU GAG GAG	CCU G**U**G GAG
amino acid sequence	**Pro** **Glu** **Glu**	**Pro** **Val** **Glu**
hemoglobin	HbA	HbS
RBC structure		



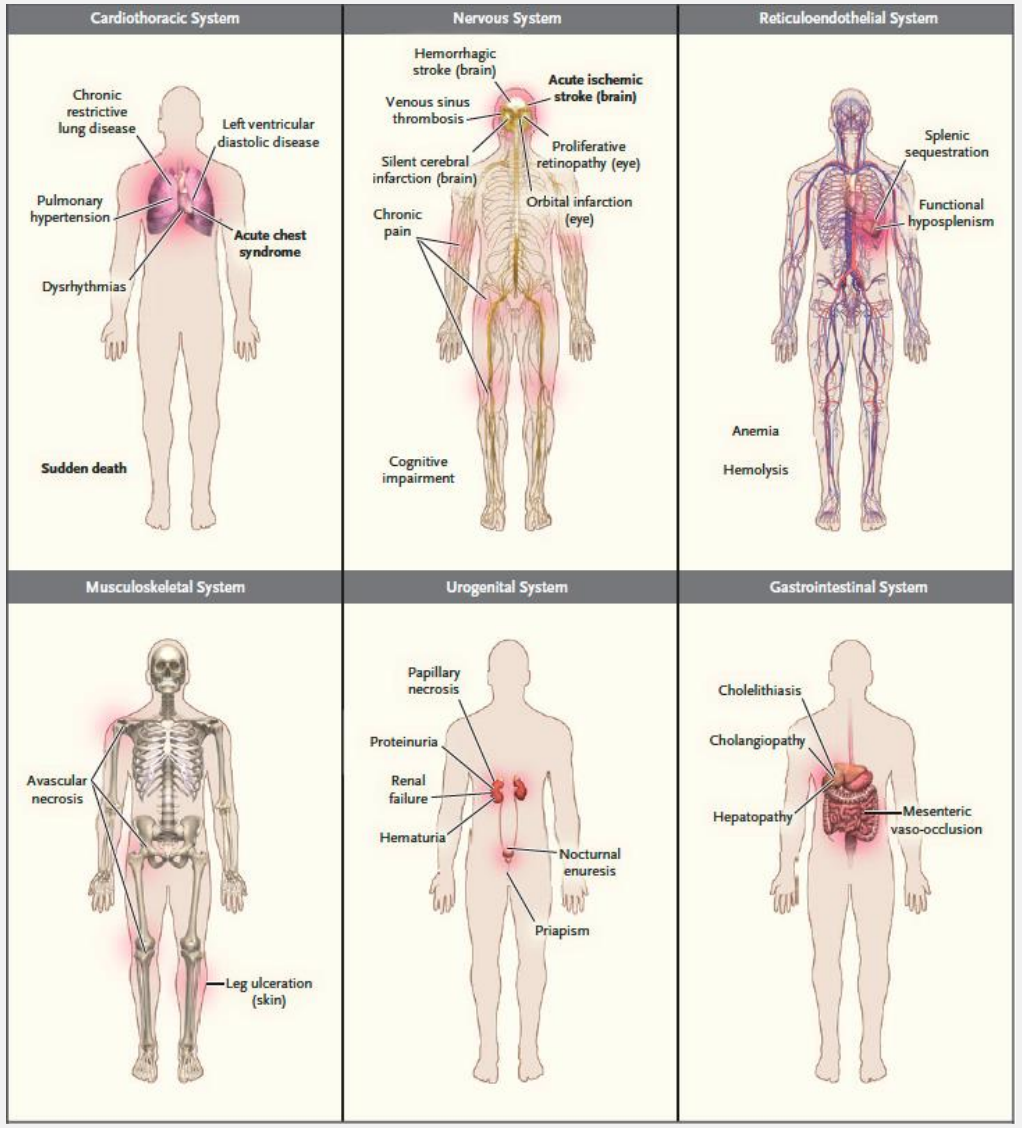




Modern distribution of malaria

Common clinical complications of sickle cell disease

1. Anemia
2. Vaso-occlusion
   - Pain episodes
   - Stroke
   - Priapism
   - Acute chest syndrome
   - Renal papillary necrosis
   - Splenic infarction
   - Leg ulcers
3. Chronic organ damage
   - Lungs
   - Kidneys
   - Gallbladder
   - Eyes
   - Joints
   - Heart
   - Spleen



Piel et al., *NEJM*, 2017

Inequities and disparities in sickle cell disease

- SCD disparities and inequities mirror existing “racial”, ethnic, and economic inequities and disparities in US and globally
- Median life expectancy reduced by at least 30 years in all countries, greater in low-income countries
- Africa has highest SCD birth prevalence and mortality rate - increased mortality (50-90%) in children under age 5

NASEM, *The National Academies Press*, 2020; Tewari et al., *Haematologica*, 2015; Piel et al., *NEJM*, 2017

Inequities and disparities in sickle cell disease

- SCD has received relatively little attention and few resources from from the scientific, clinical, and public health communities compared to other genetic disorders such as cystic fibrosis.

- Funding for SCD has been historically low, compared to federal and private funding for other conditions, and has decreased over the years.

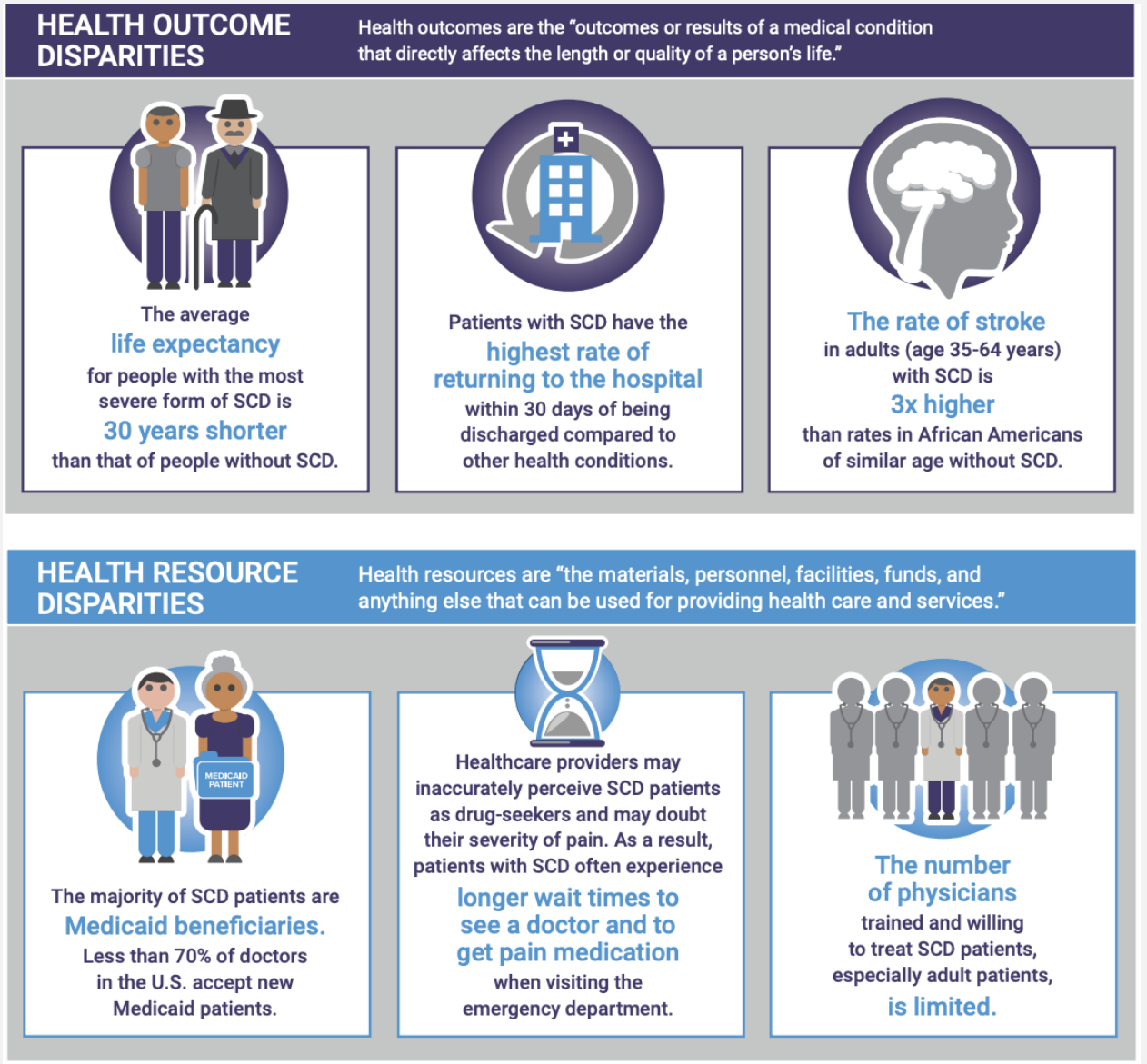
- The burden of SCD on individual patients exceeds that of numerous other severe illnesses.

NASEM, *The National Academies Press*, 2020

Sickle Cell Disease Health Disparities

"As a group, people with SCD experience worse health outcomes compared to other diseases and have access to fewer health resources."

CDC Foundation
Together our impact is greater



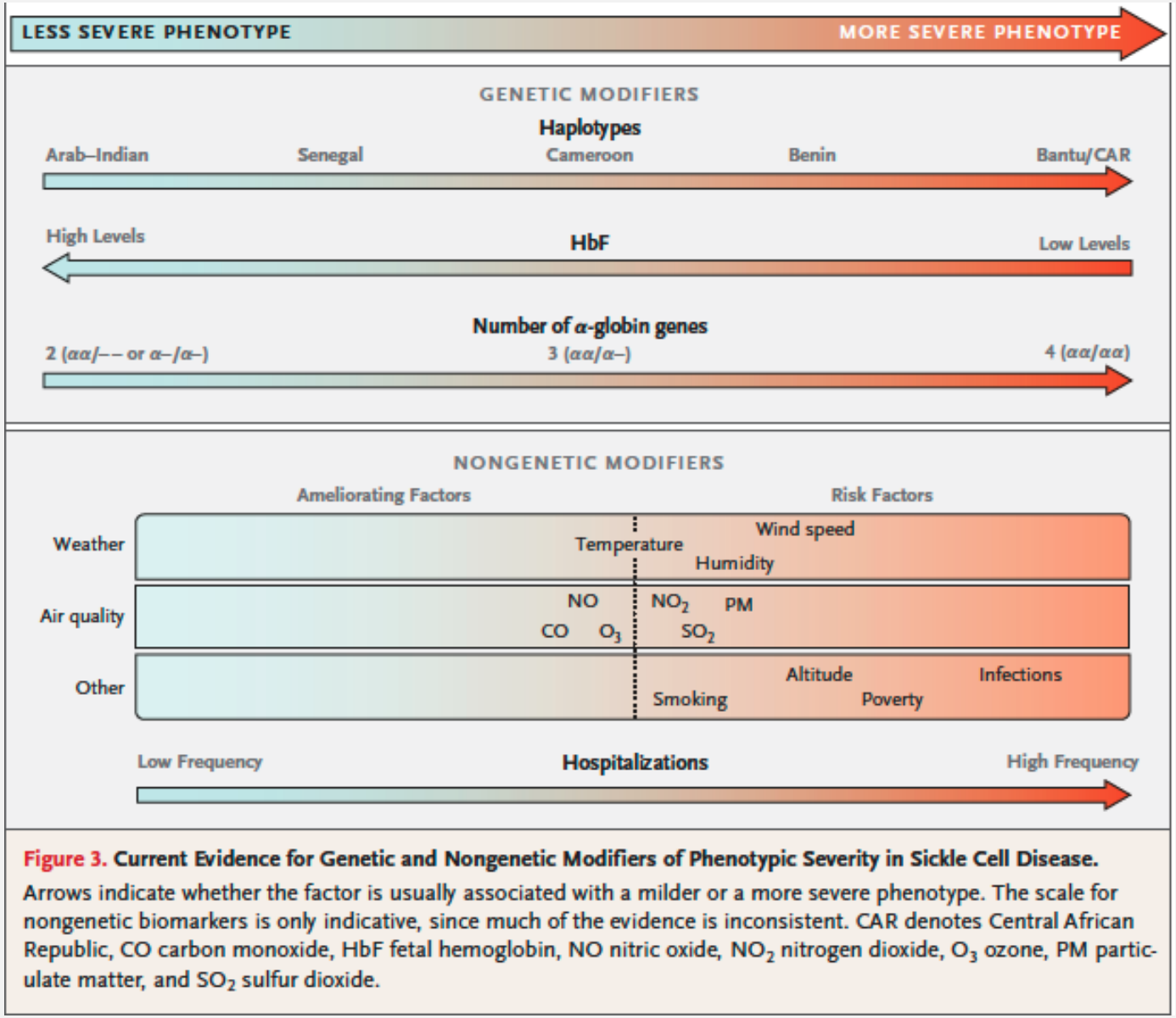


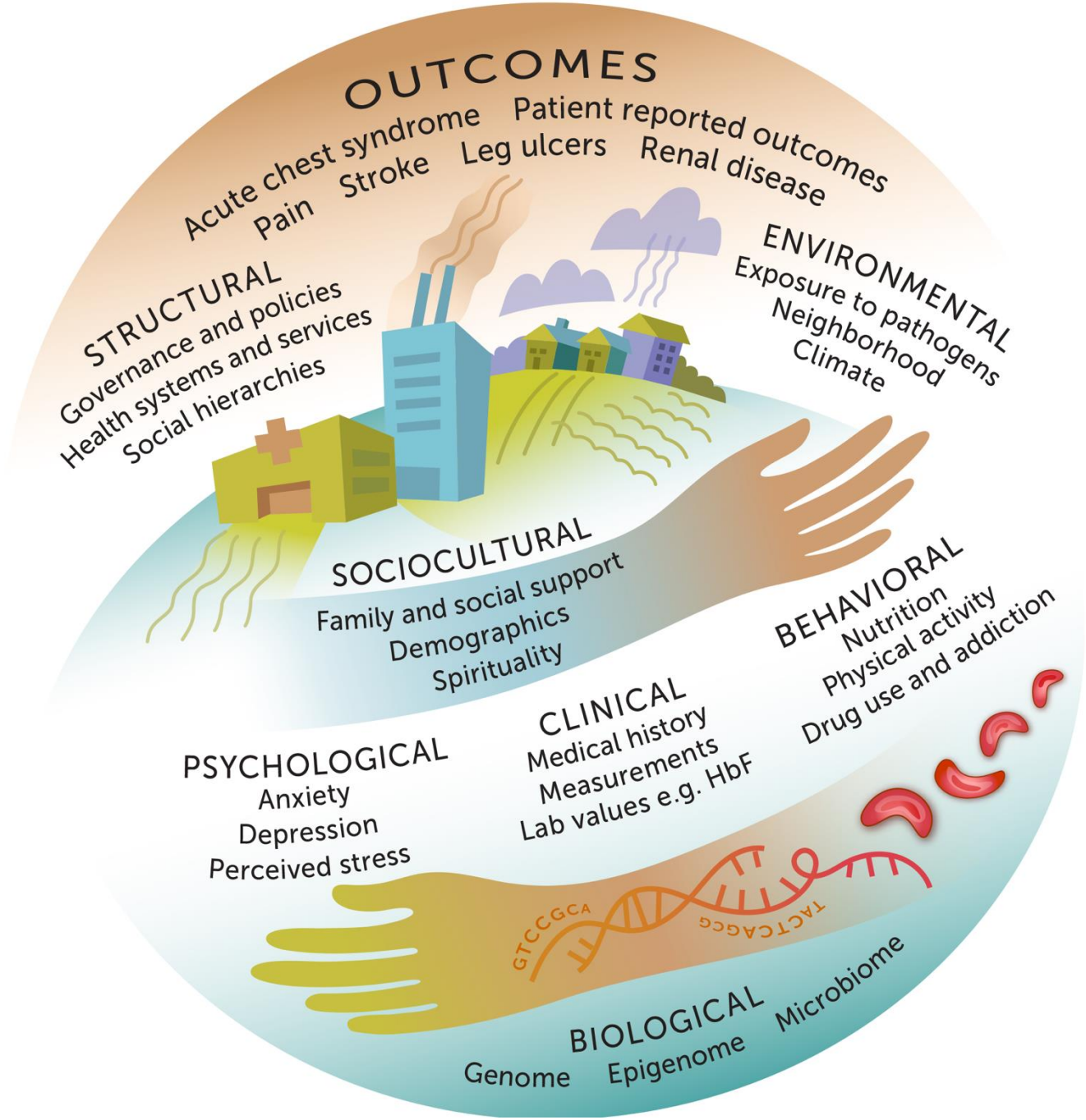
Figure 3. Current Evidence for Genetic and Nongenetic Modifiers of Phenotypic Severity in Sickle Cell Disease.

Arrows indicate whether the factor is usually associated with a milder or a more severe phenotype. The scale for nongenetic biomarkers is only indicative, since much of the evidence is inconsistent. CAR denotes Central African Republic, CO carbon monoxide, HbF fetal hemoglobin, NO nitric oxide, NO_2 nitrogen dioxide, O_3 ozone, PM particulate matter, and SO_2 sulfur dioxide.

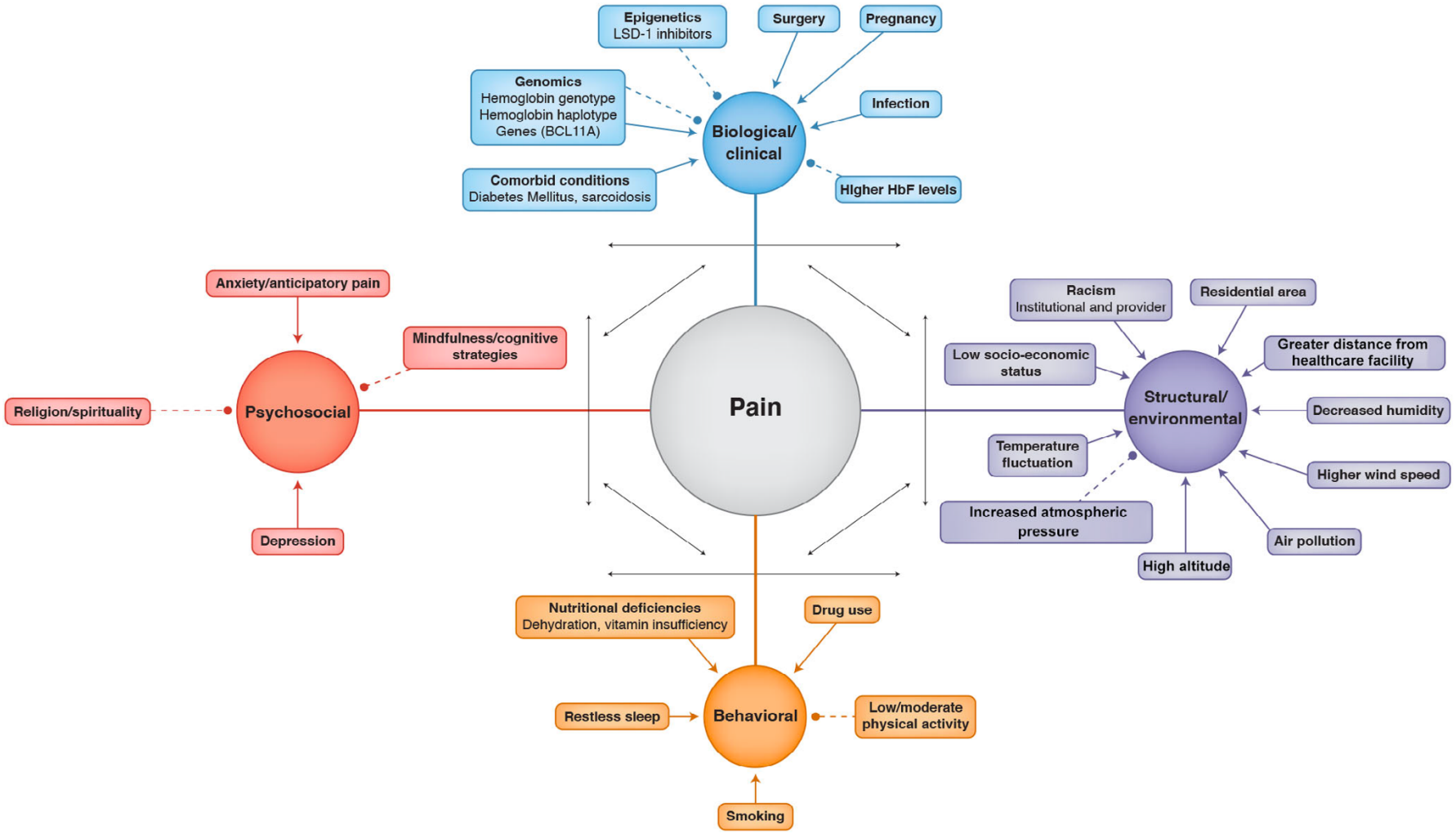
Piel et al., *NEJM*, 2017

SCD Theoretical Framework

Royal et al, *Advanced Genetics*, 2021

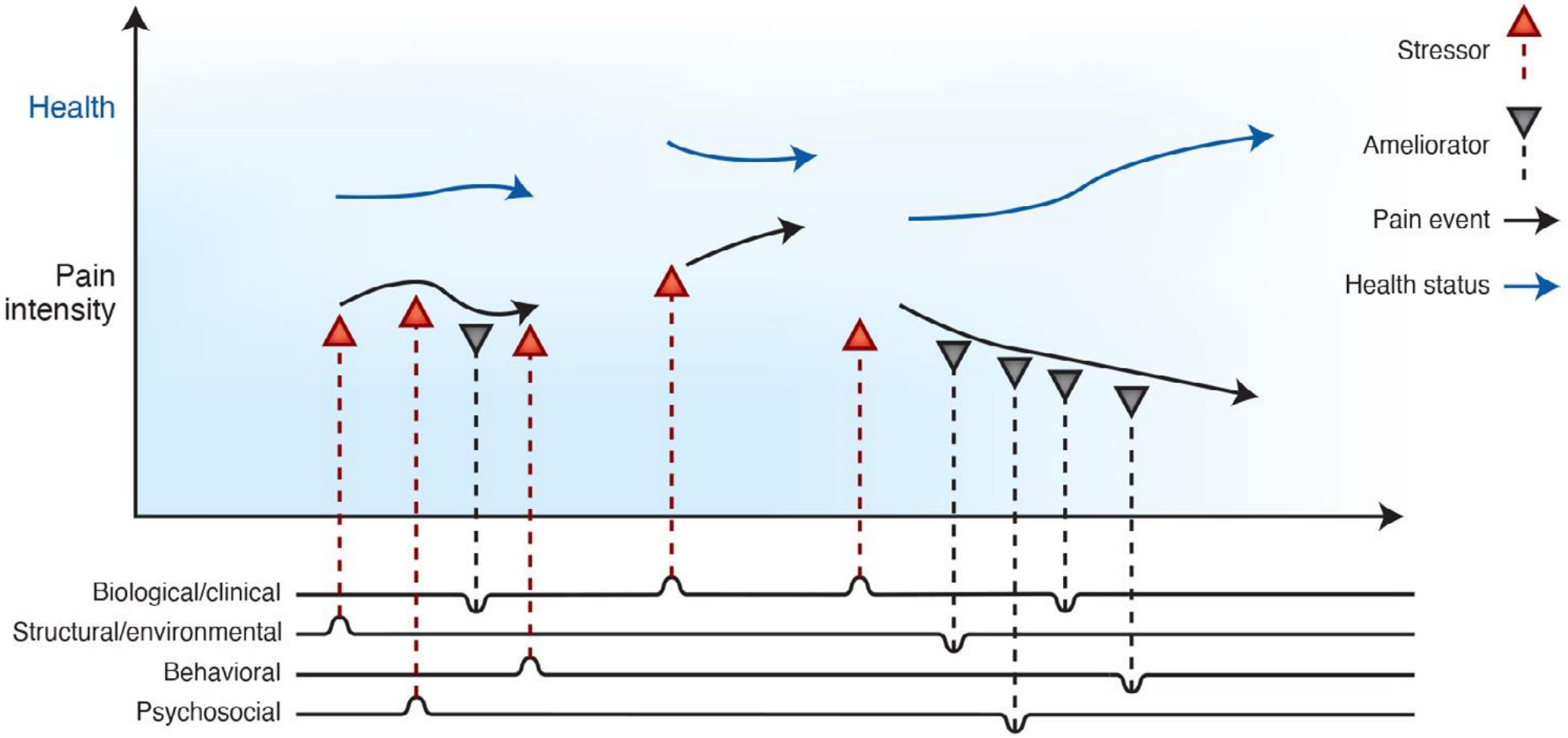


Conceptual model for sickle cell disease pain

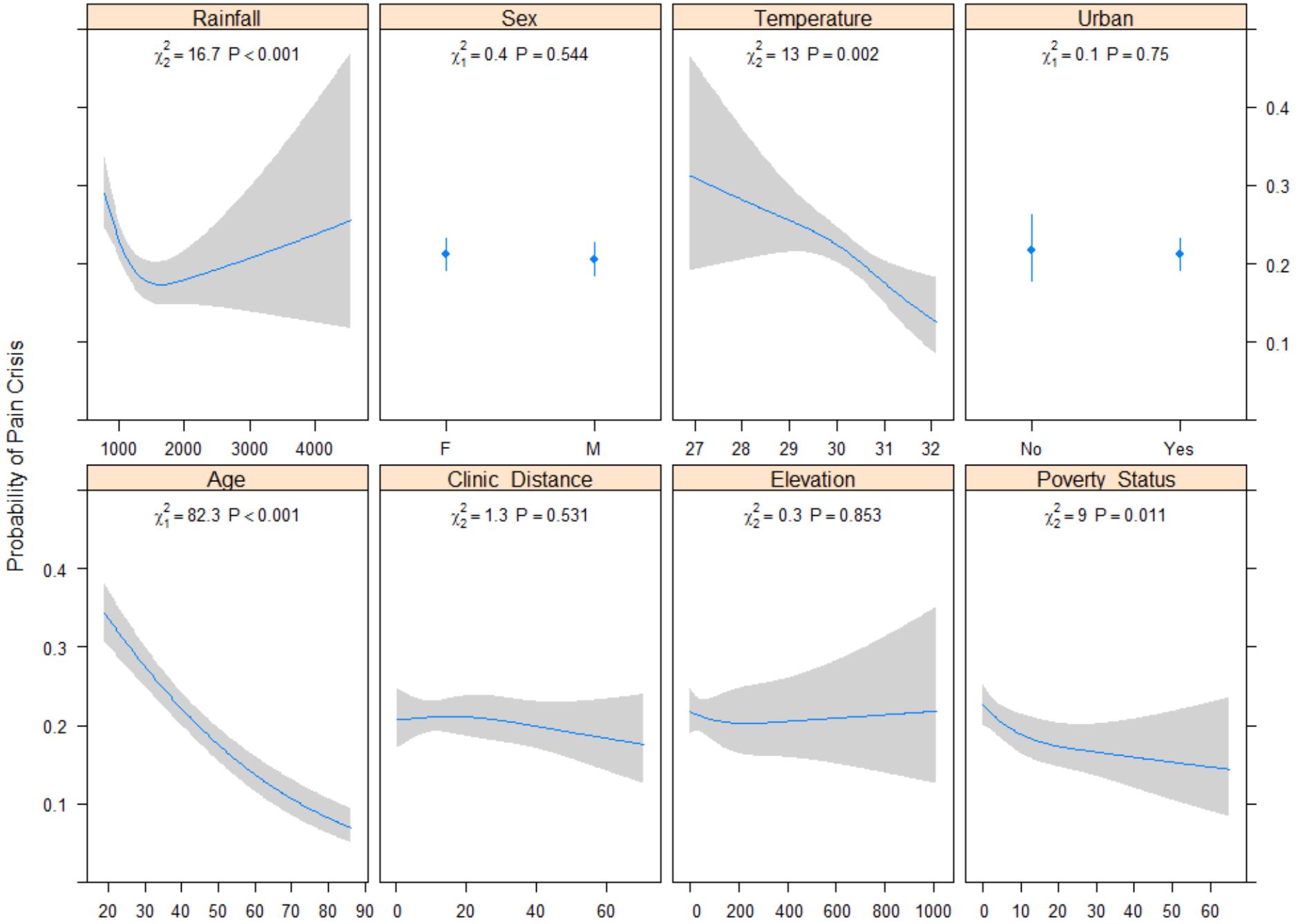


Royal et al, *Advanced Genetics*, 2021

Conceptual health timeline



Royal et al, *Advanced Genetics*, 2021



Preliminary Study: Demographic and environmental associations with painful episodes

Acknowledgements

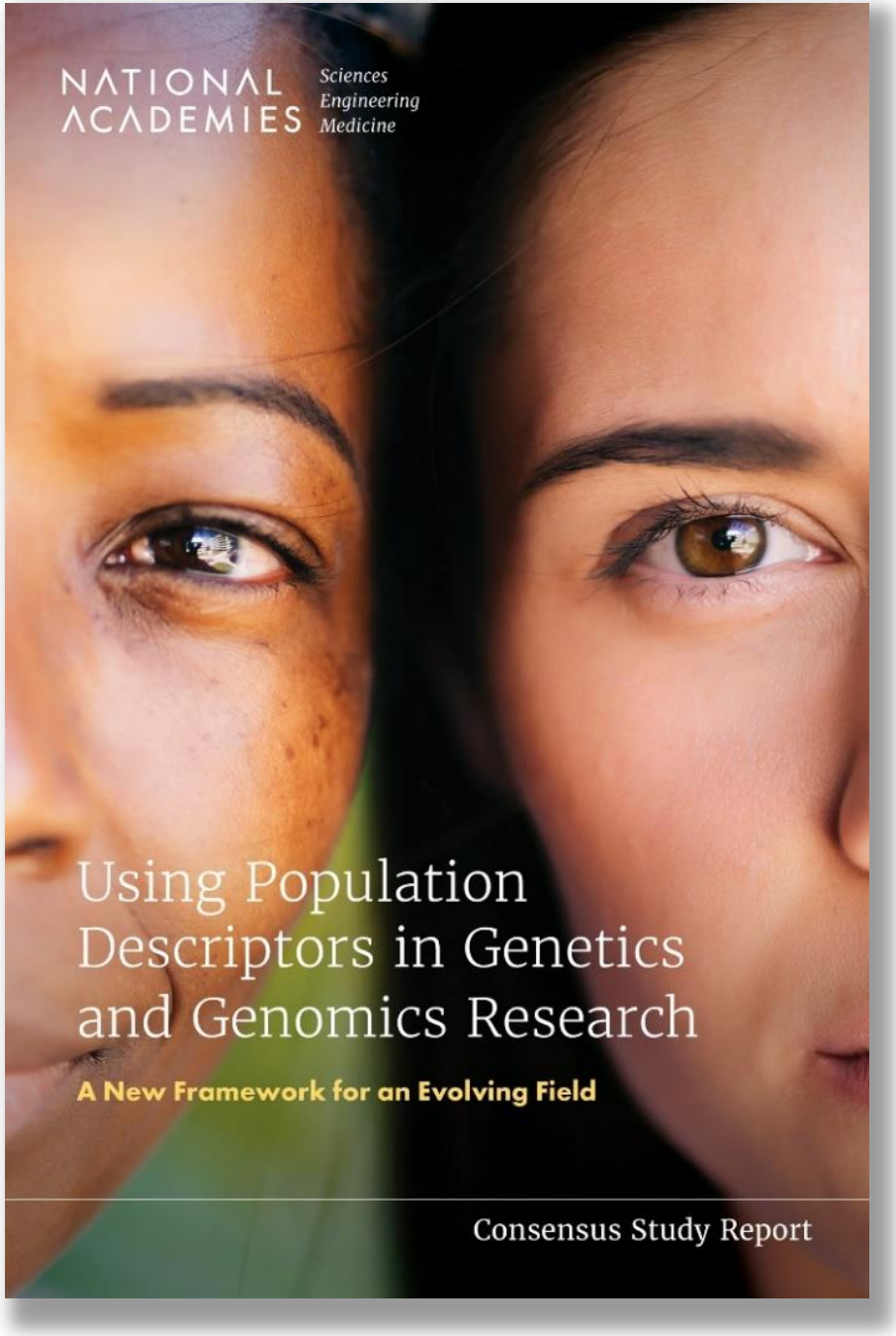
Sarah Beachy, Ph.D.
Study Director, NASEM

Samantha Schumm, Ph.D.
Associate Program Officer, NASEM

The Study Committee

- Free PDF of report is at http://www.nap.edu/
- Related materials are available through the study page, including
  - Report Highlights
  - Recommendations
  - Scrolling Page with Interactive for helping investigators decide on population descriptor use in genomics studies

NATIONAL ACADEMIES Sciences Engineering Medicine



Additional Report Resources

- Report Highlights
- Recommendations 3-pager
- Action Guides
- Interactive Webpage
  - decision tree
  - genetic ancestry video
- Press Release
- FAQs Section
- Audio Highlights and Recommendations
- Infographic
- Report Webinar and Briefing Slides

Committee Membership

Aravinda Chakravarti, Ph.D. (Co-Chair) *NAM/NAS*
New York University Grossman School of Medicine

Charmaine Royal, Ph.D. (Co-Chair)
Duke University

Katrina Armstrong, M.D. *NAM*
Columbia University

Michael Bamshad, M.D.
University of Washington & Seattle Children's Hospital

Luisa Borrell, Ph.D., D.D.S., M.P.H.
City University of New York, NY

Katrina Claw, Ph.D.
University of Colorado Denver – Anschutz Medical Campus

Clarence Gravlee, Ph.D.
University of Florida

Mark Hayward, Ph.D.
The University of Texas at Austin

Rick Kittles, Ph.D.
Morehouse School of Medicine

Sandra (Soo-Jin) Lee, Ph.D.
Columbia University

Andrés Moreno-Estrada, M.D., Ph.D.
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional

Ann Morning, Ph.D.
New York University

John Novembre, Ph.D.
University of Chicago

Molly Przeworski, Ph.D. *NAS*
Columbia University

Dorothy Roberts, J.D. *NAM*
University of Pennsylvania

Sarah A. Tishkoff, Ph.D. *NAM/NAS*
University of Pennsylvania

Genevieve Wojcik, Ph.D.
Johns Hopkins Bloomberg School of Public Health

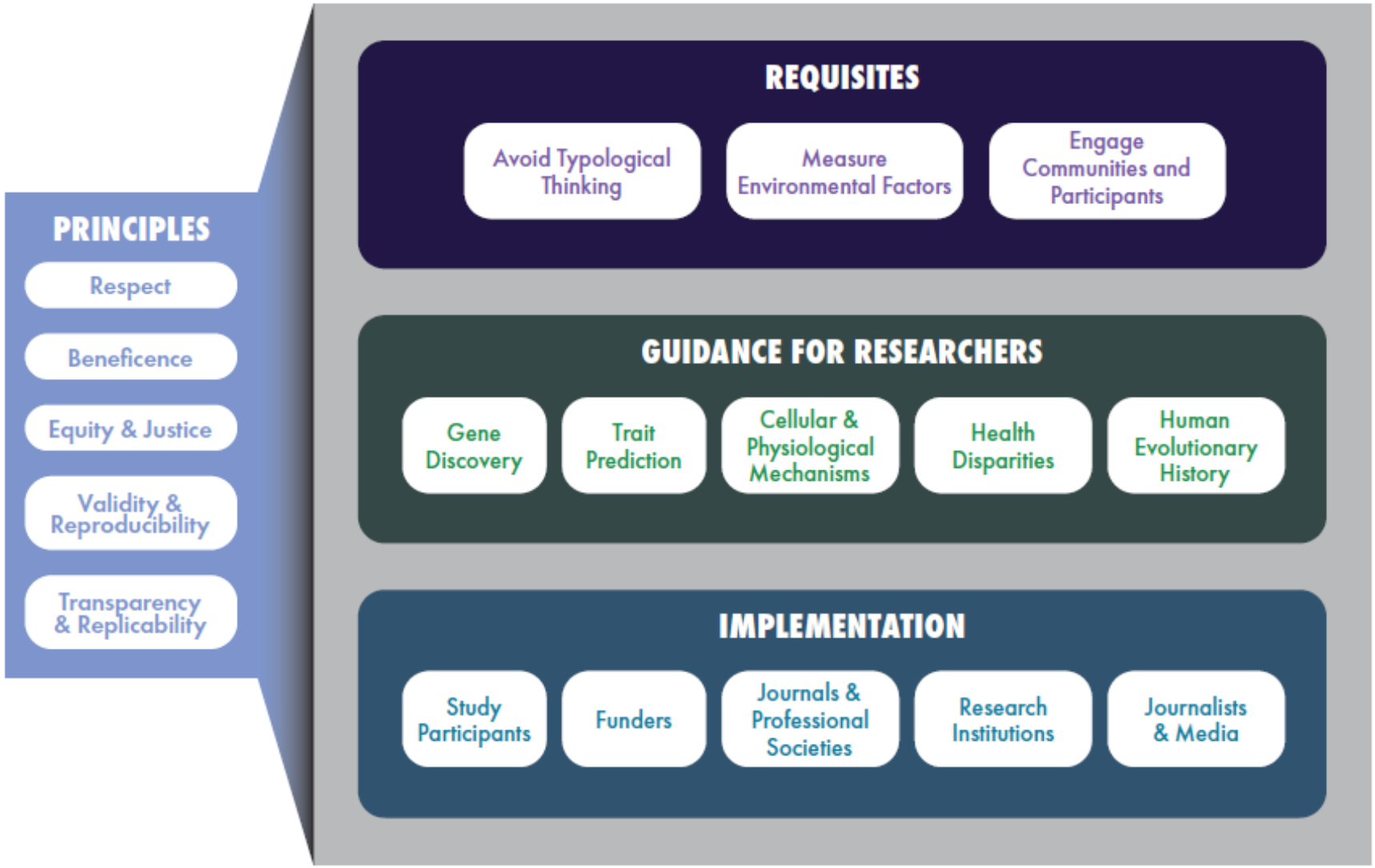
Statement of Task

- **Assessing use of race, ethnicity, and other populations descriptors** in the basic science of genetics and genomics, health risk as a function of our genomes, and health disparities
- **Developing "best practice" approaches** for the appropriate use of population descriptors
- **Discussing obstacles to adoption and implementation** of best practices
- **Proposing potential implementation strategies** to help enhance the adoption of best practices by the research community
- **Out of scope**: use of race and ethnicity in clinical care and biomedical research generally; racism in science and genomics; providing policy recommendations to NIH and government agencies

Population Descriptors Considered in the Report

Ancestry	A person's origin or descent, lineage, "roots," or heritage
Genetic ancestry	The paths through an individual's family tree by which they have inherited DNA from specific ancestors
Geography	Spatial location or geography can be measured by various indicators, such as an individual's birthplace, current place of residence, or series of previous residences
Ethnicity	Classifies human beings according to claims of shared heritage, often based on perceived cultural similarities (e.g., language, religion, foodways, dress, norms)
Indigeneity	Emphasizes a group's enduring tie to a particular geographic location as well as shared culture and traditions
Race	Classifies—and often ranks—human beings according to claims of shared ancestry based on perceived innate biological similarities

Overarching Framework



Overview of Recommendations

The committee developed 13 recommendations that fall into three categories

Requisites

- **Recommendations 1-5**
- For a general audience
- Overarching approaches important for the long-term success of this effort

Guidance for Researchers

- **Recommendations 6-8**
- 16 best practices for different types of genomics studies
- For researchers using genetics and genomics data

Implementation & Accountability

- **Recommendations 9-13**
- For selected key players in the research ecosystem
- To support researchers implementing these recommendations and best practices

Requisites to Sustain Change

Avoid typological thinking

- There is a misconception that humans can be grouped into discrete, innate biological categories
- Patterns of human genetic variation are complex
- Researchers should avoid the inaccurate assumptions of typological thinking (e.g., homogeneity of groups, hierarchy)
- **Recommendations 1-3**

Recommendation 1

Researchers should not use race as a proxy for human genetic variation. In particular, researchers should not assign genetic ancestry group labels to individuals or sets of individuals based on their race, whether self-identified or not.

Recommendation 2

When grouping people in studies of human genetic variation, researchers should avoid typological thinking, including the assumption and implication of hierarchy, homogeneity, distinct categories, and stability over time of the groups.

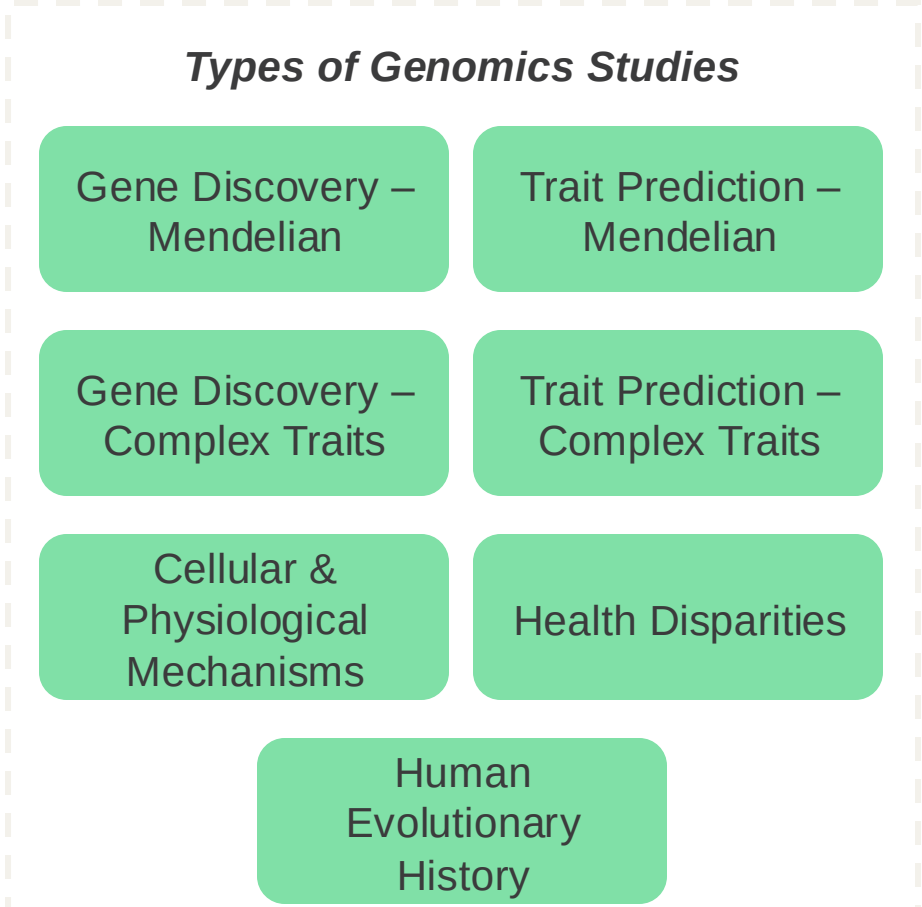
Recommendation 3

Researchers, as well as those who draw on their findings, should be attentive to the connotations and impacts of the terminology they use to label groups.

Guidance for Researchers

Researchers should tailor their use of population descriptors to the type and purpose of the study.

- There are many types of genetics and genomics studies
- There is no one-size-fits-all solution
- Researchers are decision-makers about how population descriptors are used in research. The report charges researchers to be active participants in deciding whether to use population descriptors and, if so, which ones
- Researchers should be transparent and report their decisions about population descriptors and group labels
- **Recommendations 6-8**



Examples of Guidance for Researchers

Race should not be used except for a subset of health disparities studies

Genetic similarity is a preferred descriptor in most cases

DEFINITIONS

Genetic similarity: quantitative measure of the genetic resemblance between individuals that reflects the extent of shared genetic ancestry.

Race: a sociopolitically constructed system for classifying and ranking human beings according to subjective beliefs about shared ancestry based on perceived innate biological similarities.

LEGEND

+ Preferred population descriptor(s)	− Should not be used
? In some cases; refer to Ch. 5 text and the decision tree in Appendix D	E Descriptors could be used if appropriate proxies for environmental, not genetic, effects

GENOMICS STUDY TYPE	Race	Ethnicity/ Indigeneity	Geography	Genetic Ancestry	Genetic Similarity	Notes
1: Gene Discovery - Mendelian Traits	−	?	?	?	+	Similarity suffices as a genetic measure; at fine-scale, other variables may be useful
2: Trait Prediction - Mendelian Traits	−	E	E	?	+	No population descriptors may be necessary for analysis
3: Gene Discovery - Complex Traits	−	E	E	?	+	Similarity suffices as a genetic measure
4: Trait Prediction - Complex Traits	−	E	E	?	+	Similarity suffices as a genetic measure
5: Cellular and Physiological Mechanisms	−	E	E	−	?	No population descriptors may be necessary for analysis
6: Health Disparities with Genomic Data	E	E	E	?	+	Not all health disparities studies rely on descent-associated population groupings, so none may be necessary for analysis
7: Human Evolutionary History	−	?	+	+	+	Reconstructing genetic ancestry may be of central interest

Examples of Genetic Similarity Measures

- The number of genotypes found to be identical between two individuals.
- Kinship matrices (recent genealogical ancestors)
- Similarity to reference samples (e.g., 1KG – YRI-like OR 75% of the genome is most genetically similar to individuals in the YRI panel)
- Identity-by-descent information
- Fine-scaled geographical data

Implementation & Accountability

The human genomics research ecosystem has many players that individually and collectively share responsibility for making changes and helping researchers implement the recommendations.

Study participants

Funders of genomics research

Research institutions

Journals & professional societies

Journalists & media

Recommendations 9-13

NATIONAL ACADEMIES Sciences Engineering Medicine

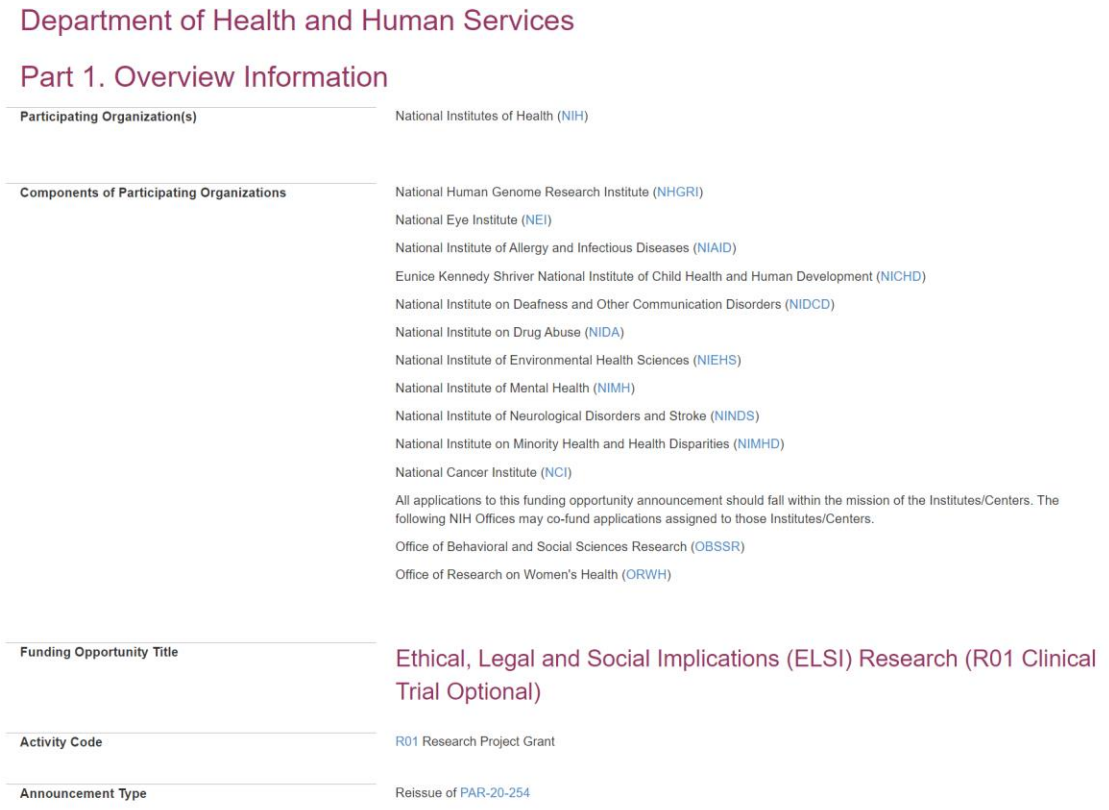
Key Points

1. The committee did not provide a menu of options, but rather a process to help researchers think through decisions about the use of population descriptors.
2. Guiding principles address ethical responsibilities and scientific standards for fostering sound best practices and trustworthy research.
3. Avoiding typological thinking, measuring environmental factors, and engaging communities are critical to achieving systemic and sustained change.
4. Genetic ancestry is inferred from various measures of genetic similarity. For many research applications, consideration of genetic similarity is sufficient without invoking the idea of genetic ancestry.
5. Use of population descriptors should depend on the nature of the study and the specific questions that the study is trying to answer. Researchers should explain how and why they decided to use the descriptors they selected.

National Institutes of Health

In October 2023, an ELSI R01 research project grant funding opportunity was announced by 11 institutes/centers and two offices with guidance on the use of population descriptors citing the NASEM report.

"Applicants who propose to address or analyze race, ethnicity, genealogical ancestry or genetic ancestry are **strongly encouraged to review the 2023 National Academies of Sciences, Engineering, and Medicine (NASEM) report**, Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field and Recommendations for Transforming the Use of Population Descriptors in Human Genetic and Genomics Research."



Department of Health and Human Services

Part 1. Overview Information

Participating Organization(s)	National Institutes of Health (NIH)
Components of Participating Organizations	National Human Genome Research Institute (NHGRI) National Eye Institute (NEI) National Institute of Allergy and Infectious Diseases (NIAID) Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) National Institute on Deafness and Other Communication Disorders (NIDCD) National Institute on Drug Abuse (NIDA) National Institute of Environmental Health Sciences (NIEHS) National Institute of Mental Health (NIMH) National Institute of Neurological Disorders and Stroke (NINDS) National Institute on Minority Health and Health Disparities (NIMHD) National Cancer Institute (NCI) All applications to this funding opportunity announcement should fall within the mission of the Institutes/Centers. The following NIH Offices may co-fund applications assigned to those Institutes/Centers. Office of Behavioral and Social Sciences Research (OBSSR) Office of Research on Women's Health (ORWH)
Funding Opportunity Title	Ethical, Legal and Social Implications (ELSI) Research (R01 Clinical Trial Optional)
Activity Code	R01 Research Project Grant
Announcement Type	Reissue of PAR-20-254

Journal Editors' Guidance

In March 2024, journal editors representing 7 biomedical journals (JAMA, Nature Genetics, American Journal of Human Genetics, Genetics in Medicine, Human Genetics and Genomics Advances, American Journal of Medical Genetics, and Journal of Genetic Counseling), published a statement providing guidance on the use of population descriptors for manuscript authors and reviewers to adopt broadly across biomedicine. This guidance was largely based on the 2023 NASEM population descriptors report.

EDITORIAL

Guidance on Use of Race, Ethnicity, and Geographic Origin as Proxies for Genetic Ancestry Groups in Biomedical Publications

W. Gregory Feero, MD, PhD; Robert D. Steiner, MD; Anne Slavotinek, MBBS, PhD; Tiago Faial, PhD; Michael J. Bamshad, MD; Jehannine Austin, PhD, CGC; Bruce R. Korf, MD, PhD; Annette Flanagin, RN, MA; Kirsten Bibbins-Domingo, PhD, MD, MAS

In March 2023, the National Academies of Sciences, Engineering, and Medicine (NASEM) released a consensus study report titled *Using Population Descriptors in Genetics and Genomics Research*.[1] Sponsored by the US National Institutes of Health, the report is more than a discussion of the use of terminology; the authors of the NASEM report suggest a tectonic shift away from current models that use race, ethnicity, and geographic origin as proxies for genetic ancestry groups (ie, a set of individuals who share more similar genetic ancestries) in genetic and genomic science. The recommendations are rooted in evidence that genetic variation in individuals falls, in general, on a continuum of variation not captured well by existing population descriptors and that the ongoing use of such descriptors as analytical variables jeopardizes the scientific validity of research.[2] Furthermore, the authors of the NASEM report point out that current scientific practices can sometimes perpetuate harmful typological thinking about individuals, including racism.

Shifting genetic and genomic science away from the pervasive and long-standing use of race, ethnicity, and geographic origins as tools for subdividing people presumed to have greater shared genetic ancestry will not be easy. The proposed changes have implications for genetic and genomic study design, data analysis, and results interpretation, and would require sustained support on the part of various stakeholders. The report offers a nuanced strategy to facilitate the shift, outlining a framework for behavior change for the field of human genetics founded on principles of respect, beneficence, equity and justice, validity and reproducibility, and transparency and replicability. These principles underlie the remaining 3 domains of the framework that include requisites for sustained change, specific guidance for the selection and use of population descriptors in genetics and genomics research, and strategies for implementation and accountability. A total of 13 recommendations are detailed in the report, each related to one of these domains. The recommendations encompass a wide variety of stakeholders in science from study participants to researchers to funders to biomedical journal editors.

Given the breadth of influence of genetic and genomic science on all areas of biomedicine, the consensus report's implications extend beyond the genetics and genomics research community to include all researchers who use genetic and genomic data as well as a broader audience. If the recommendations of the report are embraced only by genetics and genomics researchers but not more broadly, breakthrough discoveries may have scientific underpinnings that treat individuals and populations differently from how the remainder of biomedicine treats them. This could have unexpected or negative implications for the translation of genetic and genomic discoveries to the care of individuals and populations. The charge to the consensus study committee specifically excluded "examining the use of race and ethnicity in clinical care" and "examining the use of race and ethnicity in biomedical research generally (non-genetic and genomic research)", thereby focusing the report narrowly on genetic and genomics research up to the point of clinical integration.[3] The consensus report lacks concrete guidance on how to bridge potential gaps created between genetic and genomic science and the rest of biomedicine should the recommendations gain wide adoption, though further work is underway.[4] As journal editors, we believe that it is incumbent on us to help bridge any emerging gap, thereby ensuring both the scientific accuracy and interpretability of journal content.

Biomedical journals have a unique role in the translation and dissemination of genetic and genomic science to readers including researchers, clinicians, media, and the general public. The consensus report recognizes research journals as elements of the ecosystem of genomic science with a responsibility to help implement the report's recommendations.[1] Specifically, recommendation 9 suggests journals should "offer tools widely to their communities to facilitate the implementation of these recommendations," and the report includes an appendix with a checklist providing authors and reviewers guidance on the appropriate use of population descriptors in manuscripts. Recommendation 12 suggests that journals "should ensure that policies and procedures are aligned with these recommendations and invest in developing new strategies to support implementation when needed."

We journal editors concur broadly with the consensus study recommendations that population descriptors such as race, ethnicity, and geographic origin should no longer be used as proxies for genetic ancestry groups in genomic science. We also recognize that this is just one dimension of the use of population descriptors in clinically relevant research, and that drawing a distinction for requirements for genetic and genomic research and the rest of biomedicine could prove challenging. For example, the authors of the NASEM report recognize that racism can be considered a social determinant of health that can have effects on health outcomes far larger than those caused by shared genetic variation.[5]

NATIONAL ACADEMIES Sciences Engineering Medicine

NIH Workshop on Legacy Data

In May 2024, NIH hosted a meeting to discuss the NASEM report recommendations and how they relate to legacy datasets. The meeting also addressed challenges with current approaches to harmonization, interoperability and analysis, including genetic similarity and explored solutions to these issues.

All of Us Research Program
National Cancer Institute
National Human Genome Research Institute
National Institute on Aging
National Institute of Child Health and Human Development
National Institute of Diabetes and Digestive and Kidney Diseases
National Institute of Environmental Health Sciences
National Institute of Nursing Research
Office of Behavioral and Social Sciences Research
Office of Science Policy