



Beyond the Boxes, Part 1: Guiding Questions for Thoughtfully Measuring and Interpreting Race in Population Health Research

Natalie Smith, Rae Anne Martinez, Nafeesa Andrabi, Andrea (Andi) Goodwin, Rachel Wilbur, Paul Zivich

This post introduces a series our team is writing for the IAPHS blog about the use of race and ethnicity in population health research. Subsequent posts will cover our specific guiding questions on conceptualizing, measuring, and interpreting race.

Recent events in American history are motivating people and institutions to reckon with the effects of persistent racism in American society. As recent [posts](#) on the IAPHS blog have highlighted, racism still exists in the U.S., and this has been made even clearer by the COVID-19 pandemic. There has been much discussion, particularly on academic Twitter, about how researchers can better use and interpret race and ethnicity in their own work. In that vein, we wanted to call attention to an upcoming episode of the IAPHS podcast.

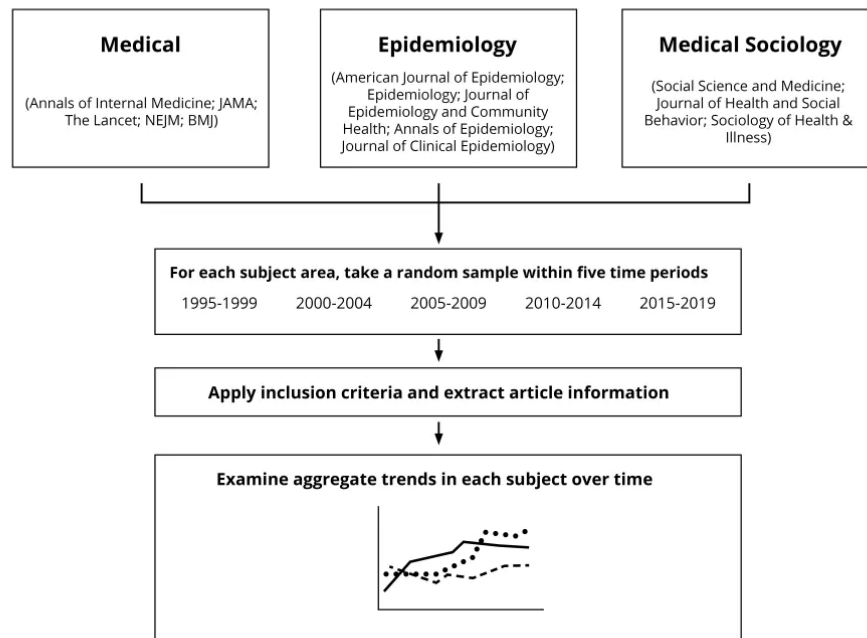
At the IAPHS 2020 Conference on October 1, 2020, at 9:00 a.m. (MDT), three members of our team (Rae Anne Martinez, Nafeesa Andrabi, and Andi Goodwin) will [discuss](#) our work on how population health researchers measure, code, and use race and ethnicity. Their discussion will be informed by our team's ongoing research project that evaluates how race and ethnicity are incorporated into population health research articles.

As students, we have been exposed to various commentaries and recommendations around measuring and using race and ethnicity. Two examples:

- In 1996, Thomas LaVeist published "[Why We Should Continue To Study Race ... But Do A Better Job.](#)" LaVeist called on researchers to more explicitly define how they conceptualize race in their work.
- This sentiment has been echoed by the International Committee of Medical Journal Editors in their manuscript preparation [recommendations](#): "Authors should define how they determined race or ethnicity and justify their relevance."

These authors and countless others have called on researchers to carefully consider how race and ethnicity fit into their work, and to interpret findings with care. But have researchers responded to these calls? Health disparities by race persist, and current events highlight the continued need to dismantle structures that perpetuate racism in the US.

We focus on separate trends in three disciplines — medicine, epidemiology, and medical sociology — to capture standard practices in how race and ethnicity are (or are not) used in different subfields of population health. Some findings that our team will discuss include trends in the use of race and ethnicity over time, how often race and ethnicity are combined into one construct, and how these variables are coded for analyses. The basic workflow of our project is shown in the figure.



After the podcast is released, we will publish additional posts in this blog series, “**Beyond the Boxes: Guiding Questions for Thoughtfully Measuring and Interpreting Race in Population Health Research.**” This series will summarize guiding questions and resources that our group of interdisciplinary trainees have used to critically reflect on how we use race and ethnicity in our own work.

We look forward to engaging with many of you during our presentation at the conference, and we hope that the subsequent blog posts will be useful resources as population health scholars everywhere work to strengthen and clarify the use of race and ethnicity in their research.

Share

Tweet

SHARE

About The Authors



Natalie Smith

Natalie is a 4th year doctoral candidate in Health Policy and Management at UNC who focuses on applying systems science, statistical, and econometric methods to public health policy questions.



Nafeesa Andrabi

Nafeesa Andrabi is a 4th year Sociology PhD student at UNC Chapel Hill, a Biosocial trainee at the Carolina Population Center and a National Science Foundation Graduate Research Fellow. Her research centers on understandings of race among Muslims in the United States and the mental and physical health consequences of Muslim racialization in the US context.



Andrea (Andi) Goodwin

Andi Goodwin is a 4th year graduate student in UNC's Sociology doctoral program, and a predoctoral trainee in the Carolina Population Center's biosocial training program. Her general research interests include social stratification's association with health disparities and caregiving burden in the US context.

Rae Anne Martinez

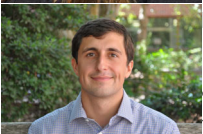


Rae Anne Martinez is currently an Epidemiology PhD student at the University of North Carolina – Chapel Hill. Martinez is broadly interested in how the built environment and social interactions are embodied through biological processes and how this embodiment impacts health across the life course. Specifically, her dissertation research focuses on the embodiment of childhood adversity via the accumulation of epigenetic alternations and links to depression and anxiety in adulthood. She is also interested in exploring historical and contemporary conceptualizations of race and ethnicity in health scholarship and the relationship of these social constructs to health. Her work is currently supported by the Biosocial NIH T32 Training Program at the Carolina Population Center, which emphasizes interdisciplinary training and collaboration in order to foster unique, innovative approaches to public health.



Rachel Wilbur

Rachel Wilbur is a fourth year doctoral candidate in the department of Anthropology at the University of North Carolina, where she previously completed an MPH in Health Behavior and Health Education. Her research explores the effects of social determinants of health, historical, and intergenerational trauma and resilience on contemporary health among Native American and Alaska Native populations. She is of Tolowa and Chetco descent.



Paul Zivich

Paul Zivich is an epidemiologist who specializes in causal inference, with a specific focus on interference, infectious diseases, and software. Paul is a PhD candidate in epidemiology at the UNC Gillings School of Global Public Health, and a Biosocial trainee at the Carolina Population Center.

◀ Member of the Month: Merlin Chowkwanyun

The World Is Your Oyster: Advice for PhD Students
Considering Non-Academic Jobs ▶

RELATED POSTS



Beyond the Boxes, Part 2: Defining Race and Ethnicity

Natalie Smith, Rae Anne Martinez, Nafeesa Andrabi, Andrea (Andi) Goodwin, Rachel Wilbur, Paul Zivich

This post is the second in a [series](#) we are writing about the use of race and ethnicity in population health research. This post covers definitional challenges and discusses why researchers must clearly state how they define race and ethnicity.

Rigorous population health research begins with a clearly stated research question, conceptual model or theory, and precisely defined exposures and outcomes. Authors commonly spend several paragraphs in the Introduction and Methods sections defining and contextualizing their exposures and outcomes. However, throughout our review of how race and ethnicity are used in population health, one of our primary insights has been that authors often do not state how they define race and ethnicity. Startlingly few articles provide explicit definitions of these concepts: of 350 randomly sampled articles from Epidemiology journals, we found **only one** that defined race and ethnicity.

Ultimately, we seem to take for granted that we know exactly what population health researchers mean when we say “I measured race and ethnicity.” We assume that our understandings of these concepts are universal across disciplines, place, and time. But that is simply not the case. Failing to define race and ethnicity makes it hard to compare findings across studies and often leaves us wondering what exactly findings pertaining to “race” or “ethnicity” really mean.

Defining the terms

Our group’s definitions are shown in the following figure and are the foundation for subsequent posts. Importantly, these definitions may be different from what you use. Including these definitions serves to underscore a key point of our posts: **it is crucial for researchers to clearly state and justify assumptions and choices around race and ethnicity in their work.**



Our team's definitions of race and ethnicity

Loosely based on definitions from Nancy Krieger's
"A Glossary for Social Epidemiology"

Race is a fluid, relational, contextually-specific social construct often derived by assigning social meaning to observable characteristics like skin tone or hair texture.

Ethnicity is a fluid, relational, and contextually-specific social construct. It is often used to reflect cultural commonalities. We've spent many hours discussing if and how ethnicity is distinct from country of origin.

Importantly, we find the essentialist perspective of race unacceptable. This perspective asserts that race is a biological construct based on immutable, inherent differences between racial groups rooted in genetics. This is an off-shoot of biological determinism which has been used to validate slavery, eugenics, genocide, unethical medical experimentation on the basis of one race being genetically superior to another. Belief in the essentialist perspective persists today, leading to continued harm & exacerbation of health disparities.

Krieger, N. (2001). A glossary for social epidemiology. *Journal of Epidemiology & Community Health*, 55(10), 693-700.

We ask ourselves the following guiding questions when **conceptualizing** race and ethnicity in our own research:

1. How do you define race? How do you define ethnicity? How are your definitions for race and ethnicity similar or different? Do they overlap?
2. Why is race and/or ethnicity important to your research question?
3. Where do race and/or ethnicity fit into your conceptual model?
4. In your manuscript, have you communicated your answers to questions #1, 2, and 3 clearly?

For us, answering these questions helped us discover conceptual gaps in our own work.

The need for establishing a shared understanding of race and ethnicity in population health research is illustrated by some challenges we've wrestled with in our work:

1. **Are Black and African American distinct identities?** The current US census form does not distinguish between these two groups. Recent work in the [New York Times](#) and discussions on Twitter (e.g., [here](#) and [here](#)) and [elsewhere](#) have dug into this question, with a variety of opinions emerging. To our knowledge, there is no one clear or correct answer. In our work, we conceptualize Black as a racial identity, while we consider African American as a US-centered ethnicity that captures a shared heritage or history, such as ancestral legacy of slavery.
2. **To whom does *Hispanic, Latino/a, Latinx* refer? Are these ethnic or racial categories?** While often used interchangeably, these US-centric terms have different foundations. Hispanic emphasizes strong connections to Spain (due to colonization), while Latino/a refers to folks from any country in Latin America currently residing in the US (and thus deemphasizes colonization, and excludes Spain). Latinx is an emergent term that first originated in LGBTQA online communities and is tightly tied to gender politics. A recent [Pew](#) report investigated perceptions of Latinx. We found it unsurprising that a highly heterogeneous group of people had varied opinions on their preferred term for ethnic identity. Hispanic, Latino/a, and Latinx are all pan-ethnic identities, and most individuals within those groups have more specific nuanced ethnic identities such as "*Colombiano/a*," "*Chicano/a*," "*Nuyorican*," or "*Tejano*." For a deeper dive into these differences, here are two readings [\(1\)](#), [\(2\)](#), or podcast options if you'd prefer to listen [\(1\)](#), [\(2\)](#).
3. **How do we best conceptualize Indigenous identity?** Indigenous peoples are often relegated to the "other" category in large data sets. When they are included, it is typically as a blanket category. Common terms for Indigenous peoples of the unceded territory of the continental United States and Alaska include Native American, American Indian, and Alaska Native. Much like Hispanic and Latino terminology, these blanket terms can be problematic because there are more than 574 federally recognized tribes in the United States, each with distinct membership requirements, histories, and culture. Still more tribes are recognized by states or remain unrecognized by government entities. It may be more appropriate to speak of individual tribal communities as ethnic groups. For an example of how three tribal nations prefer to be identified, see [the 2020 Cherokee Scholars' Statement on Sovereignty and Identity](#). For a more in-depth exploration of Indigenous race and ethnicity, we suggest reading Kim Tallbear's book *Native American DNA*.

There are no easy answers to the questions we've mentioned above, and this is just a small snapshot of the complexities of race and ethnicity that population health researchers must address. As researchers, simply knowing or acknowledging that race and ethnicity are socially constructed is not enough. We must continuously engage with the unique historical, social, and political contexts in which these categories have emerged and

continue to evolve. These historical and contemporary processes of constructing boundaries around race and ethnicity are central to understanding health disparities in the US. We would be remiss to gloss over the tensions and contradictions that exist within contemporary meanings of race and ethnicity in favor of clear, reductive boundaries

So we call on all health scholars to define your terms and look into the conceptual gaps these definitions may bring up, because these are the necessary first steps for all of us to do better in our population health research.

Our next posts will present guiding questions and resources related to measurement, analysis, and interpretation of race and ethnicity findings.

To explore these topics further, [listen to the IAPHS podcast "Sick Individuals, Sick Populations."](#)

Resources:

1. TallBear, K. (2013). *Native American DNA: Tribal belonging and the false promise of genetic science*. U of Minnesota Press.
2. Krieger, N. (2001). A glossary for social epidemiology. *Journal of Epidemiology & Community Health*, 55(10), 693-700.

Share

Tweet

SHARE

About The Authors

Natalie Smith

Natalie is a 4th year doctoral candidate in Health Policy and Management at UNC who focuses on applying systems science, statistical, and econometric methods to public health policy questions.

Nafeesa Andrabi

Nafeesa Andrabi is a 4th year Sociology PhD student at UNC Chapel Hill, a Biosocial trainee at the Carolina Population Center and a National Science Foundation Graduate Research Fellow. Her research centers on understandings of race among Muslims in the United States and the mental and physical health consequences of Muslim racialization in the US context.

Andrea (Andi) Goodwin

Andi Goodwin is a 4th year graduate student in UNC's Sociology doctoral program, and a predoctoral trainee in the Carolina Population Center's biosocial training program. Her general research interests include social stratification's association with health disparities and caregiving burden in the US context.

Rae Anne Martinez

Rae Anne Martinez is currently an Epidemiology PhD student at the University of North Carolina – Chapel Hill. Martinez is broadly interested in how the built environment and social interactions are embodied through biological processes and how this embodiment impacts health across the life course. Specifically, her dissertation research focuses on the embodiment of childhood adversity via the accumulation of epigenetic alternations and links to depression and anxiety in adulthood. She is also interested in exploring historical and contemporary conceptualizations of race and ethnicity in health scholarship and the relationship of these social constructs to health. Her work is currently supported by the Biosocial NIH T32 Training Program at the Carolina Population Center, which emphasizes interdisciplinary training and collaboration in order to foster unique, innovative approaches to public health.

Rachel Wilbur

Rachel Wilbur is a fourth year doctoral candidate in the department of Anthropology at the University of North Carolina, where she previously completed an MPH in Health Behavior and Health Education. Her research explores the effects of social determinants of health, historical, and intergenerational trauma and resilience on contemporary health among Native American and Alaska Native populations. She is of Tolowa and Chetco descent.

Paul Zivich

Paul Zivich is an epidemiologist who specializes in causal inference, with a specific focus on interference, infectious diseases, and software. Paul is a PhD candidate in epidemiology at the UNC Gillings School of Global Public Health, and a Biosocial trainee at the Carolina Population Center.

◀ Member of the Month: Jennifer Dowd

Beyond the Boxes, Part 3: Measuring Race and Ethnicity to Align with Your Research ▶

RELATED POSTS

FORUM ROUND-UP SUMMARY

Population Health News Roundup: January 2021

FORUM RESEARCH METHOD & DATA

Beyond the Boxes, Part 6: Final Thoughts and Gratitude to Our Communities

FORUM RESEARCH METHOD & DATA

In-Person Research During COVID-19: Considerations and a Call to Action

Leave a Reply

Your email address will not be published. Required fields are marked *

COMMENT

NAME *

EMAIL *

SAVE MY NAME, EMAIL, AND WEBSITE IN THIS BROWSER FOR THE NEXT TIME I COMMENT.

POST COMMENT



Beyond the Boxes, Part 3: Measuring Race and Ethnicity to Align with Your Research

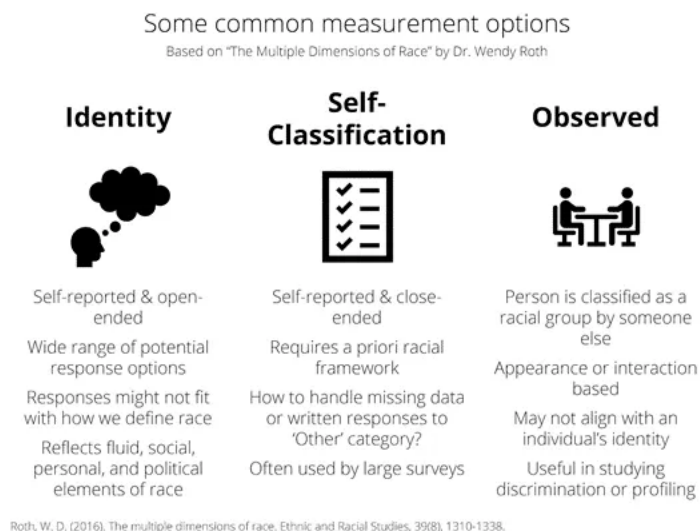
Natalie Smith, Rae Anne Martinez, Nafeesa Andrabi, Andrea (Andi) Goodwin, Rachel Wilbur, Paul Zivich

This post is the third in a [series](#) we are writing about the use of race and ethnicity in population health research. Our [second post](#) covered the challenges of defining race and ethnicity. This post presents guiding questions you can use to measure race and ethnicity in ways that align with your stated research question and conceptual model.

“...Salvador, a restaurant worker in New York, identifies his race as **Puerto Rican**. Phenotypically, he is dark-skinned with Indigenous features, leading some Americans to view him as **Black**. He believes that Americans view him as **Hispanic**, based on his accent and name. Yet on the census, Salvador checks **White** for his race because no listed option fits his identity and in Puerto Rico his mixed racial ancestry allowed him to consider himself closer to White than to Black.” ([Roth, 2016, The multiple dimensions of race](#))

Over the past few months, we’ve examined how population health researchers have measured race and ethnicity in their studies. To do this, we collected information on how race was measured using [dimensions](#) proposed by [Dr. Wendy Roth](#) as a guideline. Roth asserts that race is a fluid, multifaceted identity that can be assessed through numerous dimensions. Each dimension captures slightly different, and sometimes conflicting, information about race. Some example dimensions include racial identity, racial self-classification, and observed race (see Figure below).

Overall—across disciplines and time—our investigation found that a vast majority of studies simply do not include enough information for us to know how race or ethnicity was measured. In other words, most of the time the measure of race or ethnicity was “not stated or unclear.” As researchers, the lack of description and clarity were disappointing.



Click to enlarge.

In the above figure, we highlight dimensions of race commonly used in population health that researchers may find useful in their own work. Here's some more explanation on these terms.

Racial Identity is a self-reported, open-ended measure. When researchers ask for someone's racial identity, we might receive responses that don't match how *we* would define race because racial identity is fluid, personal, and political. Our understandings of race and ethnicity vary by person, place, and time. However, we often impose our own assumptions onto others when we clean and code racial identity data.

For example, when working with secondary data, how would you recode individuals who responded "Puerto Rican" or "Iranian" when asked an open-ended question about their racial identity? If you only want to receive information based on your own assumed racial categories (i.e., racial self-classification), or you don't want to analyze a wide range of racial and ethnic identities that may not fit with your definition, then racial identity is probably not the dimension of race you want to measure! This is exactly why we argue that it is critical to clearly [define](#) race and ethnicity, and state their relevance to your research question, before you collect your own data or analyze secondary data.

Racial Self-Classification is a self-reported, close-ended measure where individuals are asked to identify themselves within a given racial schema. Respondents might experience cognitive dissonance if they don't clearly see themselves fitting within that schema. Some may reject the racial schema all together. When we consider missing data for racial self-classification, we should ask ourselves, "Is this person's data missing because they fundamentally do not see themselves within these categories?"

Many self-classification questions have the option of "Some other race: ____" where respondents are asked to fill in their race if none of the given categories apply. Is this question really racial self-classification, or is it actually an identity/self-classification hybrid (i.e., is it the same as respondents filling in an open-ended question under "Some other race")? What kind of information are we collecting, and is that information consistent across individuals, groups, and contexts?

Observed Race is captured when an individual is assigned a racial identity by someone else. Individuals may "read" the total sum of another's body—their skin tone, hair texture, eye color, bone structure, visual cultural cues, and so on—to ascribe a racial category. Think about when a health care professional assigns a patient's race at intake in the emergency room without asking the patient how they identify. In addition to visual cues, an individual may "read" interactions—accent, gesturing, body language, and so on—to assign individuals into a perceived racial category. Think about a telephone-based survey where an individual is assigned a racial category based on accent, vocabulary, and cadence alone. Especially in studies that rely on healthcare data, we may be analyzing observed race without even knowing it (e.g., electronic health records).

Purposeful measurement of race and ethnicity is important because *measurement is indelibly linked to results*. Without a proper understanding of how studies have measured race and ethnicity, how can we synthesize findings across studies and draw overarching conclusions about racial and ethnic inequality in the US?

Our team's guiding questions for measuring race and ethnicity are:

1. What dimension(s) of race or ethnicity is(are) most relevant to your research question?
2. If you are collecting your own data, what kind of question(s) will you use to measure your chosen dimension(s) of race and ethnicity?
3. If your work uses existing data, how does that data source conceptualize and measure race and ethnicity?
 - a. Is it consistent with your research question?
 - b. If it is not, what are the limitations or resultant biases?
4. Measurement-specific questions to consider, specifically around self-report questions, include:
 - a. Is the self-report question open- or close-ended?
 - b. If close-ended, what is the list of potential response options? These racial and ethnic groups are the product of social forces, conscious individual decisions, and [federal reporting requirements](#) (themselves based on social forces and individual decisions).
 - c. Are participants asked what group they MOSTLY consider themselves? Or should they be asked to report on ALL of the groups they consider themselves to be?
5. In your manuscript, have you unambiguously communicated what measure(s) you used?

Given how important large datasets are to population health research, our group has been working on a living document outlining how large population-wide datasets measure race and ethnicity. We've made this document [available to view](#). This form contains the datasets that our team has used in the past, but there are many more datasets out there. We've created a [short form](#) for researchers to fill out with other datasets so that we can expand this resource. Relatedly, the [Multiple Components of Race Data Library](#) contains information on large datasets that measure race and ethnicity in multiple ways.

The next post in our series will discuss our guiding questions around the use of race and ethnicity in analyses.

Resources

1. Roth, Wendy D. (2016) The multiple dimensions of race, *Ethnic and Racial Studies*, 39:8, 1310-1338, DOI: [10.1080/01419870.2016.1140793](https://doi.org/10.1080/01419870.2016.1140793)
2. [Multiple components of race data library](#)
3. [Our resource](#) outlining how race/ethnicity are measured in existing pop health datasets. Please add other datasets using [this form link](#).
4. OMB measurement rule: [Federal Register, 1997](#); slide on OMB [Standards](#)

Share

Tweet

SHARE

About The Authors

Natalie Smith

Natalie is a 4th year doctoral candidate in Health Policy and Management at UNC who focuses on applying systems science, statistical, and econometric methods to public health policy questions.

Rae Anne Martinez

Rae Anne Martinez is currently an Epidemiology PhD student at the University of North Carolina – Chapel Hill. Martinez is broadly interested in how the built environment and social interactions are embodied through biological processes and how this embodiment impacts health across the life course. Specifically, her dissertation research focuses on the embodiment of childhood adversity via the accumulation of epigenetic alternations and links to depression and anxiety in adulthood. She is also interested in exploring historical and contemporary conceptualizations of race and ethnicity in health scholarship and the relationship of these social constructs to health. Her work is currently supported by the Biosocial NIH T32 Training Program at the Carolina Population Center, which emphasizes interdisciplinary training and collaboration in order to foster unique, innovative approaches to public health.

Nafeesa Andrabi

Nafeesa Andrabi is a 4th year Sociology PhD student at UNC Chapel Hill, a Biosocial trainee at the Carolina Population Center and a National Science Foundation Graduate Research Fellow. Her research centers on understandings of race among Muslims in the United States and the mental and physical health consequences of Muslim racialization in the US context.

Andrea (Andi) Goodwin

Andi Goodwin is a 4th year graduate student in UNC's Sociology doctoral program, and a predoctoral trainee in the Carolina Population Center's biosocial training program. Her general research interests include social stratification's association with health disparities and caregiving burden in the US context.

Rachel Wilbur

Rachel Wilbur is a fourth year doctoral candidate in the department of Anthropology at the University of North Carolina, where she previously completed an MPH in Health Behavior and Health Education. Her research explores the effects of social determinants of health, historical, and intergenerational trauma and resilience on contemporary health among Native American and Alaska Native populations. She is of Tolowa and Chetco descent.

Paul Zivich

Paul Zivich is an epidemiologist who specializes in causal inference, with a specific focus on interference, infectious diseases, and software. Paul is a PhD candidate in epidemiology at the UNC Gillings School of Global Public Health, and a Biosocial trainee at the Carolina Population Center.

[◀ Beyond the Boxes, Part 2: Defining Race and Ethnicity](#)

[Mentoring Spotlight: Shared Backgrounds Help Build a Successful Mentorship ▶](#)

RELATED POSTS

FORUM ROUND-UP SUMMARY

[Population Health News Roundup: January 2021](#)

FORUM RESEARCH METHOD & DATA

[Beyond the Boxes, Part 6: Final Thoughts and Gratitude to Our Communities](#)

FORUM RESEARCH METHOD & DATA

[In-Person Research During COVID-19: Considerations and a Call to Action](#)

Leave a Reply

Your email address will not be published. Required fields are marked *



Beyond the Boxes, Part 4: Complications in Coding Race and Ethnicity

Natalie Smith, Rae Anne Martinez, Nafeesa Andrabi, Andrea (Andi) Goodwin, Rachel Wilbur, Paul Zivich

This post is the fourth in a [series](#) about the use of race and ethnicity in population health research. Our previous posts have outlined our thoughts around [conceptualizing](#) and [measuring](#) race and ethnicity. Generally, researchers will then have to code those race and ethnicity measurements for use in analyses. This post details how we approach coding these variables.

As we've discussed, issues of race and ethnicity measurement are directly linked to analysis and interpretation. But there is another layer of complexity here: coding. The practice of coding variables—collapsing groups together or merging different variables—can fundamentally alter the results of data analysis, and ultimately, the interpretations of those results. As a part of our [larger project](#) studying how population researchers incorporate race and ethnicity into their work, we examined how population health studies code race and ethnicity. You can see some of our results [here](#). In sum, we find:

- Emphasis on binary coding schemes oriented around whiteness (i.e. “white,” “non-white”)
- Broad use of “white,” “Black,” “Hispanic,” and “other,” where “Hispanic” is used as a de-facto racial category, and everyone else is aggregated into the ambiguous “other”
- Slight variations on the above

These findings drive our guiding questions:

1. How will you code race or ethnicity in statistical analyses?
2. Have you collapsed race and ethnicity into an ethno-racial construct?
3. If you aggregate different racial or ethnic groups, what are the potential implications for your findings?

4. In your manuscript, have you communicated which groups you collapsed together, why these decisions were made, and what implications they have for interpreting your findings (biases, limitations)?

The underlying assumption when we engage with race and ethnicity variables is that these groups are meaningful in terms of history, privilege, access to resources, cultural similarities, and so on. When we collapse groups together, we implicitly make decisions about whose history, power, privilege, and so on are more or less similar and important. Sometimes, depending on the context and specific study question, collapsing might be advisable. Racial or ethnic groups may share similar histories along various axes. But oftentimes they do not. In such instances, collapsing groups is not advisable, particularly in the context of the highly ambiguous “other” category.

We are still considering how to approach the aggregation of groups—and really, we should all push ourselves to think about this more. How can we meaningfully interpret a coefficient that represents people with vastly different contexts and backgrounds? Why did we bother including a group at all if we can't produce findings with respect to their unique lived experience?

We also wanted to mention the differences between using an ethno-racial construct compared to using individual ethnicity and race constructs. These two approaches to coding have very different embedded assumptions. When race and ethnicity are kept separate for analyses, we assume that each captures distinct information, and could be related to population health outcomes in distinct ways. Conversely, when they are combined into an ethno-racial construct (e.g., non-Hispanic white, non-Hispanic Black, Hispanic, other), this assumes that race and ethnicity are capturing similar information and have similar relationships to health outcomes.

As researchers, we must be aware of the differences and similarities between race and ethnicity as constructs. Your research question might demand an ethnoracial perspective, or perhaps you're dealing with data limitations.

Some additional guiding questions on this point:

1. Did you intentionally end up with an ethnoracial construct?
2. Do you agree with the assumptions behind this position?

Racial boundaries and identities may overlap with ethnic boundaries and identities. Clear delineation can be challenging. On that note, we believe it is important to name those we did not see represented, discussed, or acknowledged in the studies we sampled.

We never saw MENA (Middle Eastern or North African), Black or Afro-Latinos, or Indigenous Latinos. We also very rarely saw the diversity of Native American/Alaskan Native, Asian, or Hawaiian & Pacific Islanders explored in health scholarship. These groups are frequently relegated to the “other” category. For example, when we treat “Hispanic” or “Latino/a/x” as a racial group, Afro-Latinos and Indigenous Latinos typically fall into the same group, masking potentially important differences between those groups. Most authors fail to state who even constitutes the “other” category. As such, we are unable to tell if folks are present in our studies but masked by coding practices, or if they aren't included at all. Regardless, more work is needed to explore the unique health concerns for these erased populations.

It is critical that we justify why we chose a particular action and that we understand the assumptions those actions entail. Over the last few posts, we have highlighted that our [conceptualization](#) of race ethnicity and how we [measure](#) these social constructs influence the assumptions we make when we collapse groups together. To be clear: We are all making assumptions that influence our methodological choices. We recommend that authors make their assumptions and choices clear in their scientific communication.

Our next post will discuss our guiding questions around interpretation of race and ethnicity in analyses.

Share

Tweet

SHARE

About The Authors

Natalie Smith

Natalie is a 4th year doctoral candidate in Health Policy and Management at UNC who focuses on applying systems science, statistical, and econometric methods to public health policy questions.

Rae Anne Martinez

Rae Anne Martinez is currently an Epidemiology PhD student at the University of North Carolina – Chapel Hill. Martinez is broadly interested in how the built environment and social interactions are embodied through biological processes and how this embodiment impacts health across the life course. Specifically, her dissertation research focuses on the embodiment of childhood adversity via the accumulation of epigenetic alternations and links to depression and anxiety in adulthood. She is also interested in exploring historical and contemporary conceptualizations of race and ethnicity in health scholarship and the relationship of these social constructs to health. Her work is currently supported by the Biosocial NIH T32 Training Program at the Carolina Population Center, which emphasizes interdisciplinary training and collaboration in order to foster unique, innovative approaches to public health.

Nafeesa Andrabi

Nafeesa Andrabi is a 4th year Sociology PhD student at UNC Chapel Hill, a Biosocial trainee at the Carolina Population Center and a National Science Foundation Graduate Research Fellow. Her research centers on understandings of race among Muslims in the United States and the mental and physical health consequences of Muslim racialization in the US context.

Andrea (Andi) Goodwin

Andi Goodwin is a 4th year graduate student in UNC's Sociology doctoral program, and a predoctoral trainee in the Carolina Population Center's biosocial training program. Her general research interests include social stratification's association with health disparities and caregiving burden in the US context.

Rachel Wilbur

Rachel Wilbur is a fourth year doctoral candidate in the department of Anthropology at the University of North Carolina, where she previously completed an MPH in Health Behavior and Health Education. Her research explores the effects of social determinants of health, historical, and intergenerational trauma and resilience on contemporary health among Native American and Alaska Native populations. She is of Tolowa and Chetco descent.

Paul Zivich

Paul Zivich is an epidemiologist who specializes in causal inference, with a specific focus on interference, infectious diseases, and software. Paul is a PhD candidate in epidemiology at the UNC Gillings School of Global Public Health, and a Biosocial trainee at the Carolina Population Center.

◀◀ [Mentoring Spotlight: Shared Backgrounds Help Build a Successful Mentorship](#)

[Population Health News Roundup: November](#) ▶▶

RELATED POSTS



Beyond the Boxes, Part 5: Analysis and Interpretation of Race and Ethnicity

Natalie Smith, Rae Anne Martinez, Nafeesa Andrabi, Andrea (Andi) Goodwin, Rachel Wilbur, Paul Zivich

This post is the fifth in a [series](#) about the use of race and ethnicity in population health research. Our previous posts have reviewed [defining](#), [measuring](#), and [coding](#) race and ethnicity. Researchers then typically move into analysis and interpretation, which is the focus of this post.

By the time we arrive at analyzing data, researchers should have (1) explicit working definitions of race and ethnicity, (2) a clear understanding of our measures, (3) a clear rationale for coding variables, and (4) an ability to justify our choices. Without giving serious attention to all of these aspects, our analyses and interpretations can at best be superficial, and at worst result in conclusions that do not reflect reality and could harm the populations we are studying.

Considerations for analyses

Descriptive aims work to describe a defined population (e.g. the proportion of individuals with characteristic “X”). These types of aims are often exploratory and could benefit from including multiple measurement schemes for race and ethnicity.

For example, a study may consider measuring both self-identification and self-classification. Open-ended measures may be preferred over requiring respondents to select from predetermined categories. Self-classification could also be measured to align with reporting requirements or later analytic methods. Directly measuring self-classification is preferred over attempting to re-code self-identity into higher aggregated categories, because that way, we are not imposing our own assumptions over others’ identities.

Predictive aims focus on creating algorithms or rule-sets to predict future characteristics. These aims often create an algorithm using training data and are then deployed to a real-world setting. In that case, studies should ensure that the training data uses race and ethnicity dimensions that are consistent with the real-world setting data. However, we note that the utility of inclusion of race or ethnicity in clinical screening tools and other predictive algorithms (e.g. prison sentencing) is [questionable](#).

Inferential aims focus on the estimation of causal effects, such as what the risk of death would have been had everyone been exposed to chemical “A.” These studies require explicit conceptualization, measurement, and coding of race and ethnicity.

As a confounder, researchers should consider what dimensions of race and ethnicity might be most relevant. Even when race or ethnicity is a confounder, it is critical to use the correct [dimension](#). For example, melanin expression, rather than self-reported race, would likely be a confounder of the relationship between sunscreen use and development of skin cancer. Another example is the relationship between taking “Medication B” and subsequent stress. In this context, racism may be the common cause of both not receiving “Medication B” and elevated levels of stress.

When race or ethnicity is of direct interest, we might be interested in the causal effect of race or ethnicity itself (perhaps as a proxy for racism). In this context, an unclear definition of “race” likely does not meet the definition of a [“well-defined”](#) exposure. Better defining race or ethnicity along specific dimensions can move us closer to a well-defined exposure, and could then be incorporated into the potential outcomes [framework](#). Another framework is to study racial or ethnic disparities through mediation analyses, as discussed [here](#) and [here](#). Finally, researchers might also be interested in the modification of another exposure effect by race or ethnicity. In that case, measures that focus on disparities between groups have also been [proposed](#).

Considerations for interpretation

In our review of epidemiology literature, we were unable to credibly assess if researchers interpreted race and ethnicity results appropriately, because as our [previous post](#) discusses, authors typically do not state how they conceptualize race or ethnicity. We offer two examples of interpretations highlighting alternative interpretations based on different conceptualizations. Consider a study where race is a “risk factor:” Where authors have stated that race is a social construct, race in a “risk factor” study might be interpreted as a proxy for structural and historical factors that influence access to resources (power, privilege, control of governmental or commercial institutions, aggregation of economic resources, and so on) and constrain life chances.

However, where authors have stated race is underlying biological and genetic variation, race in a “risk factor” study would be interpreted as immutable, innate biological differences between groups of people. Despite widespread scientific agreement that race has no biological or genetic basis, we see [time](#) and [time](#) again that scientists still interpret race as a biological difference. Focusing on interpreting “racism not race” has been highlighted in recent [months](#), and we are encouraged by this work.

As we are interpreting our analytic findings, we ask ourselves the following questions:

1. Does your interpretation align with the dimension of race that you measured?
2. Have you checked that you are not assigning a biological cause to racial or ethnic differences that were measured using a self-reported dimension?
3. Have you considered interpersonal, organizational, and structural racism as a potential explanation or mechanism?

Resources:

1. LaVeist, T. A. (2000). On the study of race, racism, and health: a shift from description to explanation. *International Journal of Health Services*, 30(1), 217-219.
2. Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *NEJM* 383:874-882.
3. Hernán, M. A. (2016). Does water kill? A call for less casual causal inferences. *Annals of epidemiology*, 26(10), 674-680.
4. Robins, J. M., & Weissman, M. B. (2016). Commentary: Counterfactual causation and streetlamps: what is to be done?. *International Journal of Epidemiology*, 45(6), 1830-1835.
5. Naimi, A. I., & Kaufman, J. S. (2015). Counterfactual theory in social epidemiology: reconciling analysis and action for the social determinants of health. *Current Epidemiology Reports*, 2(1), 52-60.
6. VanderWeele, T. J., & Robinson, W. R. (2014). On causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology (Cambridge, Mass.)*, 25(4), 473.
7. Naimi, A. I. (2016). The counterfactual implications of fundamental cause theory. *Current Epidemiology Reports*, 3(1), 92-97.

Share

Tweet

SHARE

About The Authors

Natalie Smith

Natalie is a 4th year doctoral candidate in Health Policy and Management at UNC who focuses on applying systems science, statistical, and econometric methods to public health policy questions.

Rae Anne Martinez

Rae Anne Martinez is currently an Epidemiology PhD student at the University of North Carolina – Chapel Hill. Martinez is broadly interested in how the built environment and social interactions are embodied through biological processes and how this embodiment impacts health across the life course. Specifically, her dissertation research focuses on the embodiment of childhood adversity via the accumulation of epigenetic alternations and links to depression and anxiety in adulthood. She is also interested in exploring historical and contemporary conceptualizations of race and ethnicity in health scholarship and the relationship of these social constructs to health. Her work is currently supported by the Biosocial NIH T32 Training Program at the Carolina Population Center, which emphasizes interdisciplinary training and collaboration in order to foster unique, innovative approaches to public health.

Nafeesa Andrabi

Nafeesa Andrabi is a 4th year Sociology PhD student at UNC Chapel Hill, a Biosocial trainee at the Carolina Population Center and a National Science Foundation Graduate Research Fellow. Her research centers on understandings of race among Muslims in the United States and the mental and physical health consequences of Muslim racialization in the US context.

Andrea (Andi) Goodwin

Andi Goodwin is a 4th year graduate student in UNC's Sociology doctoral program, and a predoctoral trainee in the Carolina Population Center's biosocial training program. Her general research interests include social stratification's association with health disparities and caregiving burden in the US context.

Rachel Wilbur

Rachel Wilbur is a fourth year doctoral candidate in the department of Anthropology at the University of North Carolina, where she previously completed an MPH in Health Behavior and Health Education. Her research explores the effects of social determinants of health, historical, and intergenerational trauma and resilience on contemporary health among Native American and Alaska Native populations. She is of Tolowa and Chetco descent.

Paul Zivich

Paul Zivich is an epidemiologist who specializes in causal inference, with a specific focus on interference, infectious diseases, and software. Paul is a PhD candidate in epidemiology at the UNC Gillings School of Global Public Health, and a Biosocial trainee at the Carolina Population Center.

◀◀ Population Health News Roundup: November

Member of the Month: Sean A. Valles ▶▶

RELATED POSTS



Beyond the Boxes, Part 6: Final Thoughts and Gratitude to Our Communities

Natalie Smith, Rae Anne Martinez, Nafeesa Andrabi, Andrea (Andi) Goodwin, Rachel Wilbur, Paul Zivich

This post is the final entry in our [series](#) about the use of race and ethnicity in population health research. Our previous posts have detailed information around [defining](#), [measuring](#), [coding](#), and [analyzing](#) data on race and ethnicity. A full list of our guiding questions is linked [here](#). Here we offer some concluding thoughts.

Rigorous consideration of race and ethnicity can be difficult to get completely right, but that is no excuse to stop striving to do better. The guiding questions presented in previous posts have helped our team critically reflect on our own work, and we hope that our questions, resources, thoughts, and tips will be helpful for other researchers moving forward.

These guiding questions are largely focused on how individuals conduct research. We recognize that in asking population health researchers to make their assumptions explicit, we may see a slew of work that relies on the essentialist perspective of race or other problematic assumptions. We should be prepared for this, but if we can't make implicit assumptions explicit, how can we address them?

To that end, we offer two considerations beyond individual research practices. First, we recommend [careful peer review](#). Call out essentialism and ambiguity in language when you see it. Tell editors essentialism is unfounded. Ask authors to clarify the purpose of including race and ethnicity in their study. Ask for measurement details. We also need to see fundamental [shifts in curricula](#) to move away from essentialist teachings. Thoughtful consideration of social constructs like race and ethnicity should be centered in coursework. Discussions and work surrounding these social constructs should not be viewed as “niche” or “fringe” to the study of population health; rather, they are integral to population health inquiry, and they should be celebrated as such. Furthermore, journal editors should enforce their [own standards](#) for publications.

We also want to leave readers with a few unanswered questions our team has been thinking about:

- How can we best measure structural racism and discrimination? The recent IAPHS conference had [several sessions with a focus on structural racism](#) and featured work from researchers like Dr. [Rachel Hardeman](#) and Dr. [Tyson Brown](#).
- How can we balance the need to have large enough “cell sizes” for proper estimation with the fact that data aggregation masks critical differences in lived experiences?
- Can we—and should we—work to standardize race and ethnicity measurements?
- Native Americans are unique in that tribal nations are sovereign nations. How can population health researchers work with Native communities without relying on the inappropriate racialization of Native American people? For some existing guidance on this topic, see this [IAPHS blog post](#) from Desi Rodriguez-Lonebear.
- How do we think about multiracial or multiethnic populations? Is collapsing folks into a “mixed” category the best practice? How can we better capture the diversity of lived experiences and better understand the population health of multiracial or multiethnic populations?
- Race is often treated as fixed, but race can vary for individuals [over time](#). In longitudinal analyses with multiple assessments of race, what do we do analytically?

We see time and again how discrimination and racism negatively impact the health of BIPOC (Black, Indigenous, and People of Color) in the US. As population health researchers continue to incorporate race and ethnicity into their work, we encourage them to critically review their own assumptions and actions. We still have a long way to go in order to “[do a better job](#),” but we hope you see that many of these changes can be made right now.

A document of our considerations spanning from study design to conclusions is linked [here](#). Please [reach out](#) to our study team with any feedback you might have about our ongoing work. There may be resources we have missed, and we would welcome the opportunity to expand our resource repository.

More information on our project and other resources can be found [here](#).

A final note from our team.

Our work often discusses race and ethnicity as if they are solely burdens—as if everything about our racial and ethnic identity is rooted in power, in discrimination, in pain, in oppression, in violence. This is perhaps becoming even more common in our fervor to study and dismantle systems of structural racism and disenfranchisement in health without deep, rich measures of structural and interpersonal racism.

But racial and ethnic identity are more than proxies for oppression. Here, too, are joy, love, and belonging. Within our group, we identify as Black, Indigenous, white, immigrant, Latina, Muslim, South Asian, and multiracial. We are proud of our racial and ethnic identities, because to do so is to be proud of our ancestors, of our families, and of the communities who have raised us, cherished us, and given us voice. In our identities, we find warmth, laughter, love, and comfort.

As these posts are published on the heels of Hispanic/Latinx Heritage Month and Native American and Alaskan Native Heritage Month, we want to recognize the complexities of various racial and ethnic identities, the continued struggle against oppression, as well as the joy, love, and belonging within.

Acknowledgements

We want to enthusiastically thank the many individuals who have made this work possible: Joelle Artere-Roberts, Ganga Bey, Kate LeMasters, Denise Mitchell, Brooke Staley, Dr. Bob Hummer, and Dr. Allison Aiello. We are grateful to the [Carolina Population Center](#) for training support (T32 HD091058) and for general support (P2C HD050924).

Resources

- Jee-Lyn García, J., & Sharif, M. Z. (2015). Black lives matter: a commentary on racism and public health. *American journal of public health*, 105(8), e27-e30.
- Phelan, J. C., & Link, B. G. (2015). Is racism a fundamental cause of inequalities in health?. *Annual Review of Sociology*, 41, 311-330.
- Hardeman, R. R., & Karbeah, J. M. (2020). Examining racism in health services research: A disciplinary self-critique. *Health Services Research*, 55, 777-780.

- Boyd, R. W., Lindo, E. G., Weeks, L. D., & McLemore, M. R. (2020). On racism: a new standard for publishing on racial health inequities. *Health Affairs Blog*, 10.

Share

Tweet

SHARE

About The Authors

Natalie Smith

Natalie is a 4th year doctoral candidate in Health Policy and Management at UNC who focuses on applying systems science, statistical, and econometric methods to public health policy questions.

Rae Anne Martinez

Rae Anne Martinez is currently an Epidemiology PhD student at the University of North Carolina – Chapel Hill. Martinez is broadly interested in how the built environment and social interactions are embodied through biological processes and how this embodiment impacts health across the life course. Specifically, her dissertation research focuses on the embodiment of childhood adversity via the accumulation of epigenetic alternations and links to depression and anxiety in adulthood. She is also interested in exploring historical and contemporary conceptualizations of race and ethnicity in health scholarship and the relationship of these social constructs to health. Her work is currently supported by the Biosocial NIH T32 Training Program at the Carolina Population Center, which emphasizes interdisciplinary training and collaboration in order to foster unique, innovative approaches to public health.

Nafeesa Andrabi

Nafeesa Andrabi is a 4th year Sociology PhD student at UNC Chapel Hill, a Biosocial trainee at the Carolina Population Center and a National Science Foundation Graduate Research Fellow. Her research centers on understandings of race among Muslims in the United States and the mental and physical health consequences of Muslim racialization in the US context.

Andrea (Andi) Goodwin

Andi Goodwin is a 4th year graduate student in UNC's Sociology doctoral program, and a predoctoral trainee in the Carolina Population Center's biosocial training program. Her general research interests include social stratification's association with health disparities and caregiving burden in the US context.

Rachel Wilbur

Rachel Wilbur is a fourth year doctoral candidate in the department of Anthropology at the University of North Carolina, where she previously completed an MPH in Health Behavior and Health Education. Her research explores the effects of social determinants of health, historical, and intergenerational trauma and resilience on contemporary health among Native American and Alaska Native populations. She is of Tolowa and Chetco descent.

Paul Zivich

Paul Zivich is an epidemiologist who specializes in causal inference, with a specific focus on interference, infectious diseases, and software. Paul is a PhD candidate in epidemiology at the UNC Gillings School of Global Public Health, and a Biosocial trainee at the Carolina Population Center.

◀◀ State Preemption Laws and Public Health

Member of the Month: RacismLab ▶▶

RELATED POSTS