



Ministry of
Citizens' Services

Guide on Using Categorical Race & Ethnicity Variables

For Administrative & Survey Data Users
June 2023
Version 1.1

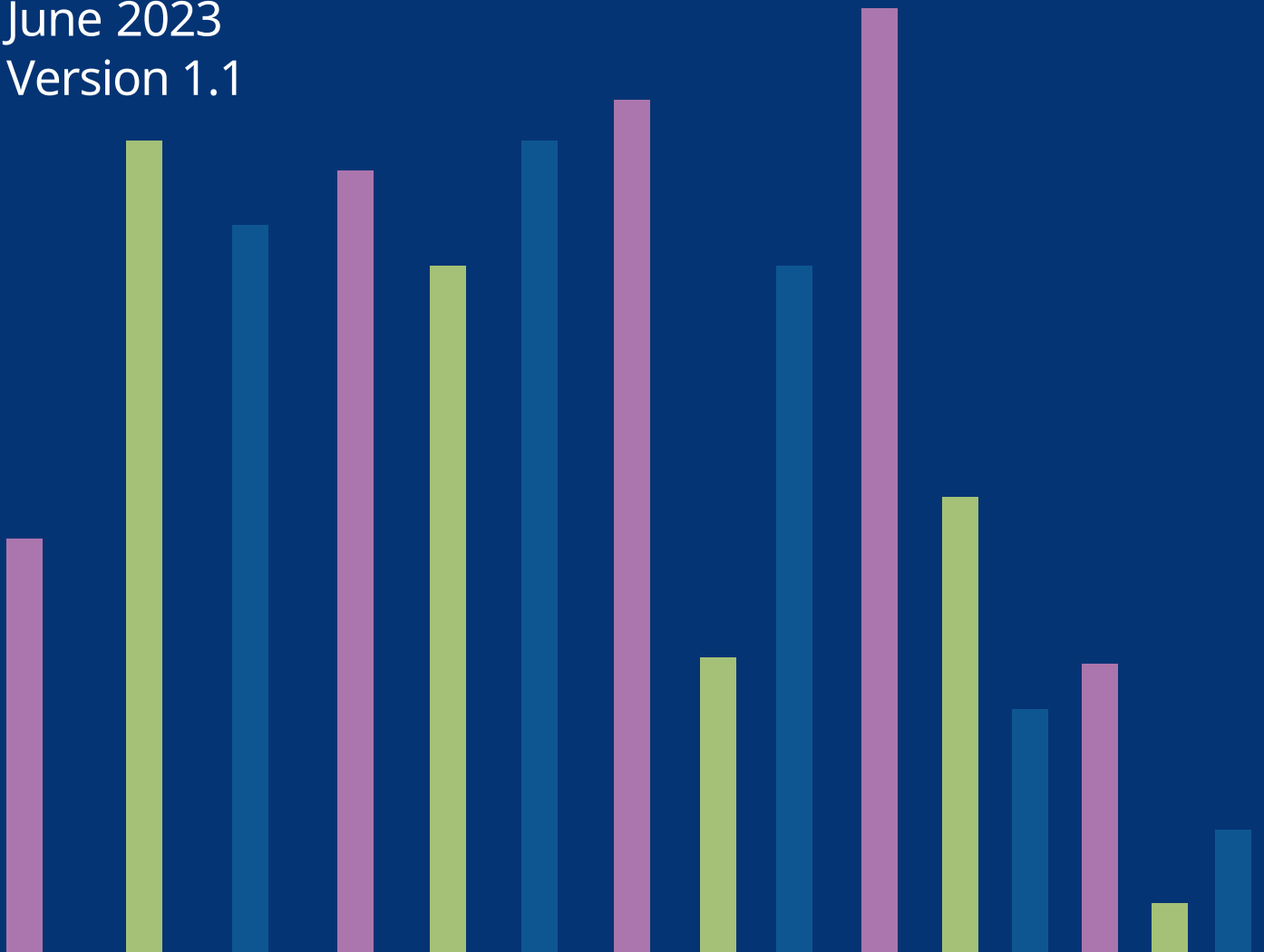


Table of Contents

Acknowledgements	1
Introduction	2
Rationale.	3
About Categorical Race and Ethnicity Variables	3
Purpose and Scope.	5
Intended Uses	6
Development Approach	7
Recommendations for Using Categorical Race and Ethnicity Variables	8
R1 Define race and ethnicity as distinct social concepts and provide a justification for their use.	9
R2 Do not attribute the effects of race and ethnicity to biological or genetic variation in racial and ethnic groups	10
R3 Consider conceptually relevant individual, social, structural, and contextual factors that explain and modify associations between race, ethnicity, and other variables. Explicitly identify what factors are included and excluded	11
R4 Describe how race and ethnicity were measured in the data used and ensure it aligns with the concept definitions used in the project.	14
R5 If grouping is necessary, ensure that resulting categorizations meaningfully represent racial and ethnic groups included	15
R6 When describing racial and ethnic groups, use terminology preferred by their members and appropriate in the context of the project	16
R7 When some racial and ethnic groups are included and others are excluded, explain why this is done	17
R8 When comparing racial and ethnic groups, explain why this is done and take steps to mitigate the risk of harm from drawn comparisons	18
R9 Include information explaining race and ethnicity-related results, their accuracy and whether they can be generalized	19

Glossary	21
Appendix 1. Race and Ethnicity as Variables	26
Appendix 2. Case Studies	27
C1 Social nature of race and ethnicity variables	27
C2 Justifying inclusion and exclusion of groups and using a medical standard for group comparisons	28
C3 Using race and ethnicity variables interchangeably.	28
C4 Inaccuracy of genetic attribution of race effects.	29
C5 Social attribution of race effects and discussion of mediators.	29
C6 Theoretical incorporation of mediators and moderators.	30
C7 Simpson’s paradox as a common error resulting from omission of moderators	30
C8 Importance of including moderators	31
C9 The complexities of comparing race and ethnicity variables from different datasets	32
C10 Data processing approaches for multiple response questions.	33
C11 The importance of considering diversity within racial and ethnic groups	33
C12 Examples of available guidelines on describing racial and ethnic groups	34
C13 Illustrating within-group variability through data visualization	35
C14 Discussing limitations to generalizability of findings that use census benchmarking.	36
Appendix 3. Self-Assessment Checklist	37
References	41

Acknowledgements

The BC Data Service Division within the Ministry of Citizens' Services (the Ministry), who work throughout the province of British Columbia, would like to acknowledge with gratitude and respect the distinct traditional territories of the First Nations peoples.

We specifically acknowledge and express our gratitude to the lək'wəŋən people of the Songhees and Xwsepsum Nations where the Ministry main office is located.

The BC Data Service Division also recognizes Métis people and Métis Chartered Communities, as well as the Inuit and urban Indigenous peoples living across the province on various traditional territories.

The BC Data Service Division also wishes to thank the external peer reviewers who contributed their time and expertise in helping create this document (in alphabetical order):

- ▶ **Fiona Clement, PhD**
Department Head and Professor,
Dept of Community Health Sciences,
University of Calgary
- ▶ **Jemal Demke, BSc**
Researcher, Wellesley Institute
- ▶ **Laura Ferreira-Legere, RN, MScN**
Senior Manager,
Public Engagement & Knowledge Translation Science Office, ICES
- ▶ **Amy Freier, PhD**
Lead, Inclusion Diversity Equity and Accessibility Team,
Health Data Research Network
Research Associate, Manitoba Centre for Health Policy,
University of Manitoba
- ▶ **Mandi Gray, PhD**
Post-Doctoral Associate, CIHR Health System Impact Fellow,
Cumming School of Medicine, University of Calgary
- ▶ **Shelly Keidar, MPH**
Project Coordinator, UBC Health,
Office of the Vice-President, Health, University of British Columbia
- ▶ **Kim McGrail, PhD**
Professor, School of Population and Public Health,
University of British Columbia
Scientific Director, Health Data Research Network Canada
(and SPOR Canadian Data Platform)
Scientific Director, Population Data BC
- ▶ **Kwame McKenzie, MD**
CEO, Wellesley Institute
Professor, Department of Psychiatry, University of Toronto
Director of Health Equity, Centre for Addiction and Mental Health
- ▶ **Arjumand Siddiqi, ScD**
Professor and Division Head of Epidemiology,
Canada Research Chair in Population Health Equity,
Dalla Lana School of Public Health, University of Toronto
- ▶ **Kienan Williams, MPH, Hons. BSc**
Program Lead, Innovation and Research,
Indigenous Wellness Core (IWC)

Introduction



Rationale



The initial appearance of **key terms** are marked with underlined bold italics and hyperlinked to their respective Glossary definitions.

Hyperlinked online sources and sections within this document are underlined and bolded.

The Province has committed to addressing all forms of *racism* and has added *anti-racism* initiatives in each ministry's [mandate letter](#). Passed in June 2022, the [Anti-Racism Data Act](#) reaffirms this commitment and helps to facilitate this work. Under the Act, the Province will use data to reveal, understand and track progress towards addressing *systemic racism* in government services and programs.

Over time, this will lead to an increased use of *race and ethnicity variables* in research and reporting projects, both inside and outside of government. Many of these projects will rely on existing administrative and survey datasets, such as those available through B.C.'s [Data Innovation Program](#), statistical programs, government records management systems and business applications. In these datasets, race and ethnicity are usually measured as *categorical variables*.¹

There are considerable differences in the way analysts, researchers, and other data users have incorporated categorical race and ethnicity *variables* into their projects. Unfortunately, instances of misuse are common for several reasons, including a lack of practical, publicly available guidance on this topic.

This guide begins to address this gap by providing recommendations on the appropriate use of categorical race and ethnicity variables in projects that rely on administrative and survey data, particularly where that data has already been collected.

In addition to this guide, the Province is working to create data standards and directives that will outline how race-based and other demographic data is collected, used and disclosed by government. The standards and directives will be made public once they are developed.

About Categorical Race and Ethnicity Variables

In this guide, race and ethnicity are understood as distinct social *concepts* or socially constructed ideas without a biological basis. These concepts categorize people into groups based on phenotypic variation and a combination of shared cultural, religious and linguistic heritage and geographic origin.

In datasets, measurable characteristics of *race and ethnicity concepts* are represented by race and ethnicity variables. Unlike abstract concepts, variables have a set of values that can be counted and manipulated for research and reporting purposes. The guide does not further define race and ethnicity concepts

¹ Categorical variables have a fixed number of values and are often used to assign people to categories or groups.

and their respective variables because the definitions will vary depending on the project and the dataset used.

The guide focuses on categorical race and ethnicity variables that represent a fixed identity descriptor because this is how they typically appear in administrative and survey data. This means, these variables capture racial or ethnic identity at one point in time and can take on a limited set of values representing racial or ethnic *categories*.

The categorical and fixed nature of these variables is emphasized in this guide to draw attention to the fact that it is not the only way race and ethnicity concepts can be measured. Racial and ethnic identity can also be seen as an evolving multidimensional aspect of self-concept and measured on a scale as a continuous variable.² Categorical variables can also capture the understanding that self-identification may change over time if measured regularly³ and the history of individual responses can be tracked.

The collective phrase “race and ethnicity” is used throughout the guide to make clear that a given recommendation applies to both variables. It does not imply that both are expected to be used together.

It is important to acknowledge that race and ethnicity concepts and their respective variables have been and still are poorly defined, used inconsistently, and misused.⁴ There are legitimate concerns about their *validity* and usefulness for research and reporting purposes.⁵

At the same time, because the concepts of race and ethnicity have become a part of our sociocultural reality⁶ and are impacting people’s lives, there are also important reasons to continue using variables that represent them. These reasons include legal reporting requirements that have been established based on these variables and the need to understand, address and prevent systemic inequities in our society.

2 For example, a Multigroup Ethnic Identity Measure developed by Phinney (1992) assesses three dimensions of ethnic identity: positive attitude and sense of belonging, identity achievement, and behaviours and practices.

3 Agadjanian (2022) illustrates how categorical survey responses can be used to capture fluidity of racial self-identification.

4 See, for example, Brown et al. (2013), Clarke et al. (2008), Comstock (2004) and Ioannidis et al. (2021).

5 See, for example, Fullilove (1998) and McKenzie & Crowcroft (1994).

6 Sociocultural reality includes the legal system, social roles and hierarchies, belief systems, physical organization of space, norms, values, language, mental models, and behaviours.

Purpose and Scope

This guide aims to facilitate research and reporting by supporting data users who work with existing administrative and survey datasets where race and ethnicity are measured as categorical variables.

✓ **The recommendations given in this guide apply to race and ethnicity variables that may appear in survey and administrative datasets under different *labels*, such as:**

- ▶ race
- ▶ ethnicity
- ▶ visible minority
- ▶ *Indigenous* identity/status
- ▶ *First Nations* identity
- ▶ *Métis* identity
- ▶ *Inuit* identity
- ▶ ethnic or cultural origin
- ▶ ethnic or cultural ancestry

✗ **The recommendations are not meant to apply to variables that measure related but different concepts, such as:**

- ▶ genetic ancestry
- ▶ language proficiency
- ▶ religious affiliation
- ▶ cultural identity
- ▶ nationality
- ▶ citizenship
- ▶ immigration status
- ▶ Registered or Treaty Indian status
- ▶ Inuit identity card
- ▶ Métis citizenship

The guide has no legal authority and is not a regulation or other legal mechanism under the Anti-Racism Data Act. These recommendations are provided simply as guidance.

While the guide summarizes a range of recommendations for research design and data analysis, interpretation and reporting, it is not intended as an exhaustive source of information on data projects involving race and ethnicity variables. The guide is not a replacement for other available training materials, professional and legal standards and ethics review and community consultation processes.

This guide assumes that the data users have an adequate level of training and are using data that has been collected and accessed through appropriate governance frameworks and consultation processes that support culturally safe use of data and prevention of community harms.⁷

Data users are responsible for understanding and respecting the rights of Indigenous Peoples regarding any data they are accessing or using (see [B.C. Declaration on the Rights of Indigenous Peoples Act](#) and the [UN Declaration on the Rights of Indigenous Peoples](#)). This includes adhering to regulations, principles, and guidelines under Indigenous data governance frameworks that uphold data sovereignty.⁸

Intended Uses

This guide consists of nine [recommendations](#) with illustrative [case studies](#) and examples.

The recommendations can be considered at planning, implementation and reporting stages of projects that use administrative and survey data. These projects include descriptive program and service reports, program evaluations and analytic studies. The recommendations are also relevant when preparing requests to access data through the B.C.'s [Data Innovation Program](#).

Not all recommendations will apply to or be feasible to implement in all projects. The recommendations may be relevant in different stages of the project and may be reviewed and considered together with project sponsors, partners and community groups.

A [self-assessment checklist](#) is included to help reflect on the extent to which the recommendations have been considered and implemented in a project.

7 The Anti-Racism Data Act defines community harm as “racism, prejudice, stereotyping, bias, stigmatization or other harm to which a group of persons is likely to be exposed”.

8 BC Office of the Human Rights Commissioner defines data sovereignty as “the right of a nation, group of people or individual to exert control over the governance of data collection, application and ownership”.

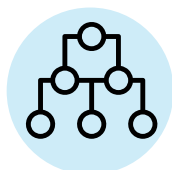
Development Approach

The guide was developed through a literature review and a peer review process which is briefly described below.



STAGE 1: Literature review

Over 200 academic articles and other literature from health and social science fields were reviewed in preparation of the recommendations and case studies for the guide. Literature that was used to develop the guide is included in the list of references.



STAGE 2: Cross-government peer review

Insights from the literature review were discussed and further refined with 60 B.C. government employees who work with administrative and survey data as part of their regular duties. These employees represented 11 (46%) of the 24 core government ministries and agencies.



STAGE 3: Peer review by subject matter experts external to government

Ten subject matter experts from B.C., Alberta, Manitoba and Ontario reviewed recommendations that emerged from the previous two stages.

[Indigenous Governing Entities](#) and the *[Anti-Racism Data Committee](#)* were given an opportunity review the guide and identify content that is culturally unsafe and has potential for community harm.

The guide is published under [sec. 19](#) of the Anti-Racism Data Act. It will be reviewed annually and updated as needed to capture additional insights that were missed in the current version and to recognize evolving practices and expectations around the use of race and ethnicity variables.

FOR MORE INFORMATION on the development process or other questions about the guide please email data-act@gov.bc.ca.

Recommendations



R1 Define race and ethnicity as distinct social concepts and provide justification for their use.



Relevant **case studies** are highlighted and hyperlinked to their location in Appendix B. Hyperlinks to navigate back to this section are also provided.

Race and ethnicity are distinct social concepts and the way they are defined, categorized and measured varies across datasets, projects, disciplines and geographies and changes over time. [SEE CASE STUDY 1](#)

Therefore, it is important to **explicitly articulate the definition(s)** used in your project and to **conceptually distinguish** race and ethnicity from one another and from related concepts (e.g., country of origin, nationality, culture, language).

Like other socio-demographic characteristics, race and ethnicity are valuable in research and reporting and can be incorporated in many ways. For the most common uses [SEE APPENDIX 1](#). When including race and ethnicity, **explain why they are relevant and what role they play in relationship to other variables**. This may be done by citing evidence supporting theoretical, policy, or program relevance of race and ethnicity. [SEE CASE STUDY 2](#)

When deciding whether to include race or ethnicity, consider if related variables are more appropriate for the purpose of your project or are preferred by the groups involved.

FOR EXAMPLE: If a project aims to understand how cultural differences relate to student achievement, it is more appropriate to use cultural identity as a variable or a combination of other variables (e.g., country of birth and age at immigration), rather than ethnicity. This is because being a certain ethnicity does not presuppose that a person has gone through the process of enculturation required to learn and adopt cultural meanings and behaviors.

Do not use race and ethnicity variables interchangeably or as a substitute for related but conceptually different measurable variables (e.g., family structure, immigration status, socioeconomic status⁹, English as an additional language status). [SEE CASE STUDY 3](#)

Similarly, do not use conceptually different variables as a substitute for race or ethnicity.

9 Socioeconomic status (SES) is a multidimensional variable that includes measures of individual or group (e.g., family) education, income and wealth. Commonly used SES measures are education level, income level, neighborhood income quintiles, home ownership and access to health insurance or high-speed internet.

R2 Do not attribute the effects of race and ethnicity to biological or genetic variation in racial and ethnic groups.

Individuals of different races and ethnicities go through race- and ethnicity-specific experiences that can have tangible positive and negative effects on their:

- ▶ biologic outcomes (e.g., physical health)
- ▶ psychologic outcomes (e.g., mental health)
- ▶ socio-economic outcomes (e.g., income)¹⁰

When modeling and interpreting these effects, do not attribute them to biological or genetic differences between racial and ethnic groups (genetic *attribution*). Instead, attribute them to the differences in social conditions encountered by these groups (social attribution). This distinction marks the difference between treating race and ethnicity as biological versus social variables.

✘ Do not use Genetic Attribution

Genetic attribution links the effects of race or ethnicity to biologic or genetic differences inherent to racial and ethnic groups. This often happens when race or ethnicity is used as a proxy for biological factors that are harder to measure or are not readily accessible (e.g., a genetic trait).¹¹

FOR EXAMPLE: A study uses genetic attribution when it explores how genetic components of race explain excess mortality in a certain racial group. This is because it implies that there are genetic differences between racial and ethnic groups that have a causal effect on the outcomes experienced by these groups.

Projects exploring or interpreting effects of race and ethnicity as innate predispositions have no validity because race and ethnicity are socially constructed concepts that are not biologically or genetically distinct. In fact, there is more genetic variation within racial and ethnic groups than between them.¹²

10 See, for example, Mustillo et al. (2004) and Paradies et al. (2015) on biologic and psychologic outcomes and Park (2021) on socio-economic outcomes.

11 See, for example, a discussion by Root (2003).

12 See, for example, Kelly et al., (2009), Khan et al., (2022) and Krieger (2005).

In rare cases where racial groups have been linked to increased risk of genetic conditions (e.g., sickle cell disease among Black [populations](#)), this link is coincidental and correlational. [SEE CASE STUDY 4](#) Not only is genetic attribution technically inaccurate, but it is also unethical and potentially harmful because it is likely to perpetuate existing stereotypes and discrimination by essentializing race and ethnicity.

✓ Use Social Attribution

Social attribution links the effects of race or ethnicity to social conditions experienced by racial and ethnic groups. These experiences can have biological consequences for racial and ethnic groups, but these consequences are a product of social interactions and not an inherent part of having a specific racial and ethnic background.

FOR EXAMPLE: A study uses social attribution when it explores how access to and experiences with health-care services explains the link between race and excess mortality. [SEE CASE STUDY 5](#)

R3 Consider conceptually relevant individual, social, structural and contextual factors that explain and modify associations between race, ethnicity and other variables. Explicitly identify what factors are included and excluded.



The initial appearance of **key terms** are marked with underlined bold italics and hyperlinked to their respective Glossary definitions.

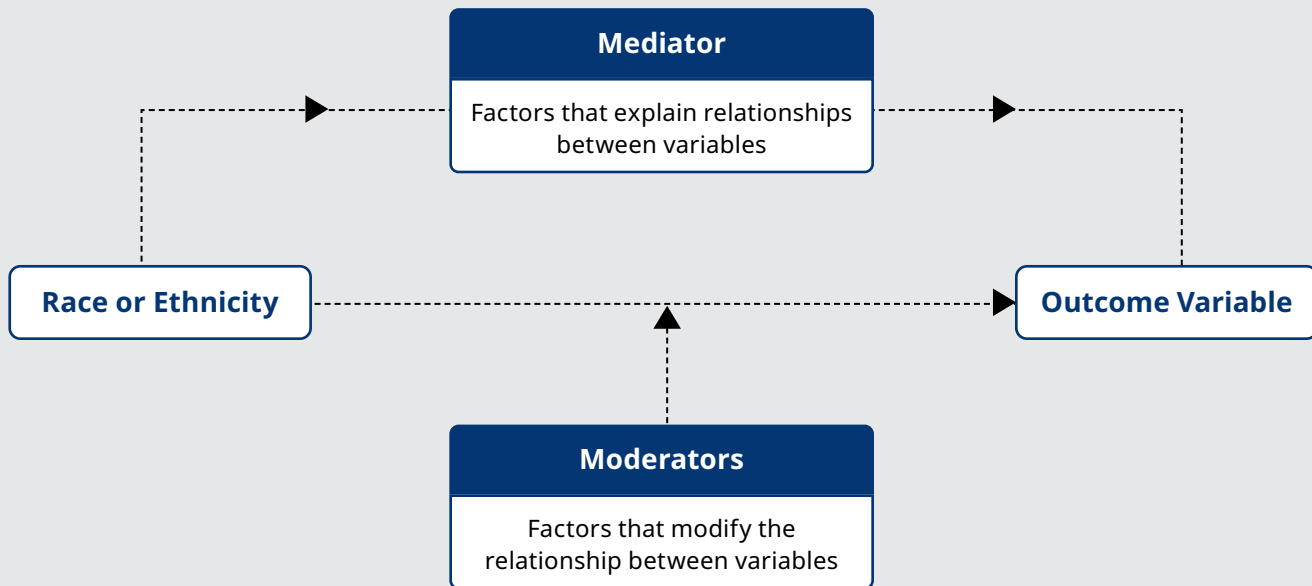
Hyperlinked online sources and sections within this document are underlined and bolded.

Since the relationship between race, ethnicity and other variables cannot be explained with biological or genetic differences inherent to these groups, it is important to consider factors that do explain why the relationship is there. These factors are called [mediators](#). It is also important to consider factors that modify the relationship between race, ethnicity and other variables. These factors are called [moderators](#) (see Fig. 1). Specifics on including mediators and moderators are discussed in sections [R3A](#) and [R3B](#).

Including mediators and moderators will increase the accuracy of findings and help develop valid interpretations of the observed relationship. Ultimately, it will help develop targeted policies and programs that direct resources to where they are needed most.

At the planning stage, consider mapping or listing relevant mediators and moderators and identifying the ones that can be included using the data available. When presenting the findings, describe what factors were not included and how this may have impacted the findings.

Fig 1. Relationship Between Variables



Information on relevant mediators and moderators can be found through a literature review and consultations with data providers, partners and racial and ethnic groups involved in the project. The same variable can be a mediator and a moderator, depending on the way it is included in the project.¹³

FOR EXAMPLE: Socioeconomic status (SES) can partially explain the link between ethnicity and post-secondary enrollment rates: ethnicity impacts one's SES, which in turn impacts one's chance of enrolling into post-secondary programs. But SES can also be modeled as a factor that modifies the link between these variables: within the same ethnicity group, individuals with higher SES may have higher post-secondary enrollment rates.

¹³ Note that mediators and moderators are different from control variables which are not discussed in this guide. Control variables are held constant in the study to remove their effects on the outcome variable of interest and their inclusion in statistical models should be carefully planned. For more information on this see, for example, Braveman (2001), Cooper & David (1986), Kaufman & Cooper (2001), Krieger et al., (1993), Morgenstern (1997) and Williams (1999).

R3A Including mediators

Mediators are variables that explain or account for the relationship between race, ethnicity and other variables. In other words, they help understand the mechanisms behind observed associations and contextualize them.

FOR EXAMPLE: Race is related to mortality, and this relationship is partially explained by access to health-care services (the mediator). It is important to include mediators because they reveal the causes of racial and ethnic inequities, encourage social attribution and discourage genetic attribution of race and ethnicity effects.

Commonly used mediators are policies and structures of social institutions, access to and quality of services and resources, family and neighbourhood characteristics, social values and norms, experiences of discrimination, effects of stereotypes and stigma, historical trauma and socioeconomic status.

Projects can incorporate mediators differently, for example by:

- ▶ including them in statistical models of analytic studies.
- ▶ providing a statistic on known mediators in descriptive reports (e.g., distance from the nearest health-care facility when reporting on health outcomes).
- ▶ offering a theoretical explanation behind the reported relationship. [SEE CASE STUDY 5](#) [SEE CASE STUDY 6](#)

At the same time, **race and ethnicity should not be completely substituted in the analysis by a set of known mediators** because the precise mechanism of their association with outcome variables is complex and may change over time.

R3B Including moderators

Moderators are variables that change the strength or direction of the relationship between race, ethnicity and other variables. In other words, they help understand if the relationship between race, ethnicity and another variable is the same across all subgroups and identify conditions that alter this relationship (e.g., protective or risk factors).

FOR EXAMPLE: A researcher may include “region” as a moderator to see if the relationship between race or ethnicity and health outcomes differs by geographical location. It is important to include conceptually relevant moderators because race and ethnicity interact with other variables (e.g., income level, age, gender) and their effects differ across population subgroups.

Including moderators will allow for a better, more nuanced understanding of the relationship between race, ethnicity and other variables and improve the precision and validity of findings. Failure to include relevant moderators may mask important within-group differences and lead to errors like those illustrated by the Simpson’s paradox. [SEE CASE STUDY 7](#)

Like mediators, moderators can be part of a statistical model or a theoretical discussion. **SEE CASE STUDY 6** In descriptive reports, moderators can be used to further disaggregate the data by various characteristics.

FOR EXAMPLE: Instead of reporting income levels by ethnicity only, report income levels by ethnicity and age group.

Commonly used moderators are:

- ▶ gender
- ▶ sex
- ▶ age
- ▶ immigration status
- ▶ education level
- ▶ income level
- ▶ geographic location

FOR EXAMPLE: Canada's history of systemic discrimination against Indigenous women makes it valuable to include gender as a moderator variable in studies about Indigenous populations. **SEE CASE STUDY 8**

R4 Describe how race and ethnicity were measured in the data used and ensure it aligns with the concept definitions used in the project.

While your project may conceptually rely on one understanding of race and ethnicity, the data available may have been collected using a different understanding of these concepts.

Ensure that the way race and ethnicity were measured in the process of data collection aligns with the concept definitions used in your project.

FOR EXAMPLE: If a project defines race as a set of physical traits, it should be using data collected to measure physical traits and not country of origin or ethnicity (even though they may be related).

If the concept definition(s) used in your project are not captured in the data available, consider if using this data is still justified and explore alternative data sources if not. When using multiple data sources, **avoid combining or comparing datasets collected under incompatible definitions and categorizations of race and ethnicity.** **SEE CASE STUDY 9**

To understand if the data available is appropriate for the project, explore how race and ethnicity were measured in the process of data collection. This involves

looking at what questions were asked, what categories were used, how they were assigned (e.g., open-ended, multiple choice) and by whom (e.g., self-report, ascribed). If the data collection method allowed for multiple responses to identify multiracial and multiethnic individuals, investigate how these responses were processed. [SEE CASE STUDY 10](#)

Explore what aspects of race and ethnicity concepts are not captured in the data available.

FOR EXAMPLE: Skin tone is one of the commonly measured aspects of race. While some datasets capture self-reported skin tone, others may capture socially perceived skin tone or a mix of both. These differences in measurement will have implications for project results because socially perceived and self-reported skin tones may differ significantly.¹⁴

R5 If grouping is necessary, ensure that resulting categorizations meaningfully represent racial and ethnic groups included.

Avoid combining multiple race and ethnicity categories, when possible. If data must be manipulated (e.g., due to small [sample](#) size, need to reduce the number of categories, align measures, or combine multiple datasets), **ensure the categories being combined are comparable even if they are labeled similarly.** [SEE CASE STUDY 9](#)

[Grouping](#) decisions should be based on the underlying similarity of characteristics or experiences of individuals being grouped and on the relevance of these characteristics or experiences to the question at hand. This means, the same grouping approach can work for some projects but not for others and across some racial and ethnic groups and not others. [SEE CASE STUDY 10](#)

If meaningful groups cannot be created, reassess if the use of race or ethnicity as a variable from the data available is still justified. Oversimplification of racial and ethnic categories increases the risk of entrenching racial bias, reinforcing assumptions that racial and ethnic groups are homogeneous, masking important differences within groups and producing uninterpretable results. [SEE CASE STUDY 11](#)

14 See, for example, Campbell & Troyer (2007), Feliciano (2016), and Garcia & Abascal (2016).

R6 When describing racial and ethnic groups, use terminology preferred by their members and appropriate in the context of the project.

As the social understanding of race and ethnicity changes, so do the socially acceptable terms used to describe different racial and ethnic groups.

While preferences and guidelines on terminology vary across disciplines, geographies and groups, there are general principles data users are advised to follow:

- ▶ Do not use stigmatizing and racist terminology and avoid labels that have a negative connotation, imply hierarchy or erase differences between ethnic and racial groups.
- ▶ Use precise descriptive language and labels that describe the underlying reason for grouping (e.g., “underrepresented in the survey” instead of “other”). When possible and relevant, describe what specific groups are included under one label.
- ▶ Use the terminology preferred by the groups in question and in line with professional, provincial or community standards and guidelines.

SEE EXAMPLES OF GUIDELINES IN CASE STUDY 12

Recognizing that data users are constrained by the labels present in the dataset, these principles may be most feasible to implement when new groupings are created and in report writing.

FOR EXAMPLE: Instead of writing “South Asians”, use a descriptive “individuals that identify as South Asian” or “South Asian students”.

While the practice of re-labeling original categories present in the dataset may be common, given the pace with which terminology evolves, the pros and cons of this practice were not explicitly discussed in the literature reviewed for this guide. On one hand, re-labeling may mask important limitations and biases built in the data collection process, make it difficult to compare results from different projects that used the same data and impact the *reliability* and validity of the findings.

FOR EXAMPLE: If racist terminology is used in the process of data collection, it may impact people's answers and, subsequently, the quality of data.

On the other hand, non-original labels may be used without distorting the original meaning of categories and to the benefit of the project.

FOR EXAMPLE: Re-labeling may be appropriate in cases where the original labels in the dataset do not reflect the actual response category used in the data collection process (e.g., the questionnaire has a "White" response option that is coded as "non-visible minority" in the dataset).

Data users are encouraged to think of the potential benefits and risks associated with re-labeling and be explicit about the decisions made. When changes to the original labels are made, they need to be documented. When changes are not made and the dataset used contains outdated labels, consider setting them off with quotation marks and explaining why these labels are used in the report.

R7 When some racial and ethnic groups are included and others are excluded, explain why this is done.

When projects focus on certain racial or ethnic groups, **the reason(s)** for including some groups and excluding other groups should be explained and justified. **SEE CASE STUDY 2** Doing so will help prevent racial and ethnic groups from being unjustly excluded from projects that may benefit them. It will also prevent some groups from being overrepresented in research and reporting simply because their data is more easily accessible.

FOR EXAMPLE: The following justifications may be provided for inclusion/exclusion of certain groups:

- ▶ The group did/did not want to be included (e.g., an Indigenous community exercising their data sovereignty rights).
- ▶ The group is/is not affected by the problem in question (e.g., race-based discrimination).
- ▶ The group is/is not the target population of a program or service (e.g., mentoring and career development services within equity programs).
- ▶ The group is/is not subject to an administrative or a legislative reporting requirement (e.g., reporting on "visible minorities" under the [Employment Equity Act](#)).

While excluding groups because of small sample size is common, it may result in a chronic exclusion of racial and ethnic groups that make up small proportions of the population. Consider if there are ways to increase sample size, such as combining multiple years, using several data sources or additional collection. If using [inferential statistical](#) analyses, consider the feasibility of alternate techniques that are better suited to small samples such as [Bayesian](#) or [non-parametric](#) methods.

R8 When comparing racial and ethnic groups, explain why this is done and take steps to mitigate the risk of harm from drawn comparisons.

While comparing racial and ethnic groups can help refute misconceptions, reveal important disparities and prompt action, it can also cause harm.

Comparative projects run the risk of reporting differences that are not there or failing to report differences that are there. Furthermore, there is risk of insufficiently or incorrectly interpreting the presence or absence of differences by the authors or by their audiences.

FOR EXAMPLE: Misinterpretation of findings often contributes to “deficit thinking” when racial or ethnic groups in question are portrayed as inherently “at risk” or in need of support and as individually or collectively responsible for the observed disparities.¹⁵ This may affect program and policy planning as well as how racial and ethnic groups are perceived in society.

Studies have demonstrated that emphasizing racial or ethnic differences can reinforce stereotypes, encourage personal attribution of responsibility (blaming the victim), exacerbate stigma, distort risk perceptions and reduce support for important policies in groups who are not considered at risk.¹⁶

While it may be impossible to fully predict and prevent these risks, several steps can be taken to mitigate them:

- ▶ Explain why the groups are being compared. **SEE ALSO R1**
- ▶ Take care not to attribute group differences to genetic causes. **SEE ALSO R2**
- ▶ Explore if between-group differences are also present within groups and to what extent by discussing moderators and/or including them in data analysis. **SEE ALSO R3**
- ▶ Explain the reasons behind the observed group differences by discussing mediators and/or including them in data analysis. **SEE ALSO R3**
- ▶ Explain what the presence or absence of differences does and does not mean. **SEE ALSO R9**
- ▶ Be explicit about the reasons why a certain referent group was used for comparison and describe the limitations of the chosen approach. While comparing to the majority/dominant group may be desirable and appropriate

15 For more on this topic see, for example, Davis, L. P., & Museus, S. D. (2019) and BC Office of the Human Rights Commissioner (2020),

16 American Psychological Association (2019), Cokley & Awad (2013), Harell & Lieberman (2021), Holder & Xiong (2022), Skinner-Dorkenoo et al. (2022), and Stephens-Dougan (2022).

for some projects,¹⁷ other projects may consider comparing groups to their own historical data, a benchmark selected by the group, a population average or an accepted standard. **SEE CASE STUDY 2**

- ▶ Consider how data visualization and presentation impacts the perception of findings. There is evidence that presentation designs that emphasize differences between groups tend to increase the risk of stereotyping and personal attribution of responsibility,¹⁸ while those that illustrate variability within groups tend to decrease these risks. **CASE STUDY 13** Consider what approach aligns best with the project purpose and with the preferences of the project partners.
- ▶ When planning the release of information about group differences, assess the risks and benefits of a public release and consider alternative ways of knowledge dissemination if the risks outweigh the benefits.

FOR EXAMPLE: Several studies on COVID-19 public awareness campaigns showed that disseminating information on racial disparities led to reduced support for safety precautions, reduced empathy for those at risk and had other negative impacts.¹⁹

R9 Include information explaining race and ethnicity-related results, their accuracy and whether they can be generalized.

To the extent possible, explain what the findings do and do not mean. Strive to include information on social, structural, and other relevant factors that may **explain** race and ethnicity effects. **SEE CASE STUDY 5**

This information can be found in the literature, through consultations with subject matter experts and participatory data analysis techniques that involve community members, individuals with lived experience and other partners. Including this information will help prevent misattribution of race and ethnicity effects and avoid reinforcing the idea that race and ethnicity are measures of biologic differences.

Provide information on the limitations and accuracy of findings. Explain whether they can be generalized across time, place and groups. Refrain from extrapolating the findings beyond the project's conceptualization of race and ethnicity and beyond the groups included in the project.

17 For example, some calls to action made by the Truth and Reconciliation Commission of Canada explicitly call to identify gaps between Indigenous and non-Indigenous communities.

18 Holder & Xiong (2022).

19 Harell & Lieberman (2021), Skinner-Dorkenoo et al. (2022), and Stephens-Dougan (2022).

FOR EXAMPLE: Educational outcomes of Black immigrant students will not necessarily be representative of educational outcomes of Nigerian immigrant students, even though many of these students may identify as Black.

Consider how different factors impact the accuracy and *generalizability* of the findings. **SEE CASE STUDY 14**

These factors may include:

- ▶ diversity between and within racial and ethnic groups
- ▶ measurement and diagnostic errors
- ▶ non-response (e.g., due to mistrust)
- ▶ sample and population size
- ▶ geographic distribution, isolation or clustering of ethnic and racial groups

While it may be more feasible to include this information in analytic studies due to their scope, descriptive projects and reports can explore ways of incorporating it in titles of tables and visuals, footnotes, attachments and interactive tooltips.

The recommendations provided in this guide will be reviewed annually and updated as needed to capture additional insights that were missed in the current version and to recognize evolving practices and expectations around the use of race and ethnicity variables.

FOR MORE INFORMATION on the development process or other questions about the guide please email data-act@gov.bc.ca.



Glossary

Glossary

TERM	DEFINITION
Anti-Racism	Anti-racism is the deliberate act of opposing racism and promoting a society that is thoughtful, inclusive and just.
Anti-Racism Data Act (ARDA)	The Anti-Racism Data Act (ARDA) became law in B.C. on June 2, 2022. The law makes it possible for the provincial government to collect, use and disclose demographic data to identify and address systemic racism in a consistent and safe way. With this information, government can improve programs and services and create a more inclusive, equitable province.
Anti-Racism Data Committee	The Anti-Racism Data Committee is established under Part 7 of the Anti-Racism Data Act (ARDA). The purpose of the committee is to support government's work to implement ARDA. The committee will also support meaningful community involvement in the development of government's collection, use and disclosure of information for the purposes of addressing systemic racism and advancing racial equity. Committee members represent a wide cross-section of racialized communities and geographic regions of B.C.
Attribution	An inference about the cause of a particular outcome or an event.
Bayesian statistics	Statistical procedures based on Bayes' theorem, which proposes a way to revise and adjust estimates of probabilities of outcomes or events using prior and new knowledge about them.
Categorical variable	A variable that can take on a limited set of qualitative values or categories.
Category (of a variable)	One of the possible values of a categorical variable that can be used to calculate frequencies.
Concept (social)	An abstract or generic idea used to make sense and talk about the world. Social concepts help us organize and communicate our understanding of interactions between people.
Control variable	A variable whose effects are "controlled for" or statistically removed because they could influence the relationship between other variables of interest.
Dependent variable	A variable that changes in response to changes in the independent variable.
Descriptive statistics	Statistical procedures used to summarize and describe a dataset.

TERM	DEFINITION
Equity (social)	Fairness and justice in how different groups of people participate in and benefit from society.
First Nations	A term used to identify one of the three populations of Indigenous Peoples within Canada, the other two being Métis and Inuit.
Generalizability	The degree to which study results can be applied to other contexts and populations.
Grouping	The act of combining individual data points or categories of a variable into groups that result in new categories.
Independent variable	A variable that produces or predicts a change in the dependent variable.
Indigenous	A collective term used to refer to First Nations, Métis and Inuit people living in what is now Canada.
Indigenous Governing Entities	An Indigenous entity that exercises governmental functions and includes but is not limited to an Indigenous governing body as defined in the Declaration on the Rights of Indigenous Peoples Act.
Inferential statistics	Statistical procedures that help understand if findings about the relationship between variables in the sample can be used to make inferences about this relationship in the population of interest.
Inuit	Indigenous people who live in the Arctic regions of what is now Canada, Greenland, United States of America and Siberia.
Label	A name (numbers or text) assigned to a variable or its categories.
Mediator	Variable that explains or accounts for the relationship between other variables. It helps to understand the mechanisms behind observed associations and contextualize them.
Métis	A person who self-identifies as Métis, is distinct from other Aboriginal peoples, is of historic Métis Nation Ancestry and who is accepted by the Métis Nation. (Métis National Council)
Moderator	Variable that changes the strength or direction of the relationship between other variables. It helps to understand if the relationship between variables is the same across all subgroups and identifies conditions that alter this relationship.

TERM	DEFINITION
Non-parametric statistics	A group of inferential procedures that do not require strict assumptions about the distribution or parameters of the data being analyzed.
Population	The entire group of people or a collection of objects or events in which the project is interested.
Race and ethnicity concepts	Abstract ideas used to categorize people into groups based on phenotypic variation (differences in their physical appearance) and a combination of shared cultural, religious and linguistic heritage and geographic origin.
Race and ethnicity variables	Variables that measure some aspect of race and ethnicity concepts.
Racism	All forms of prejudice and discrimination based on race or ethnicity.
Reliability	The degree to which study results will remain the same if repeated over time or using different samples or methods.
Sample	A subset of the population included in the project.
Systemic racism	A form of racism that is built into social and organizational structures, processes, services, programs, practices, policies and legislation.
Validity	The degree to which study results accurately represent and assess concepts and relationships of interest.
Variable	A characteristic of a concept, person, object or an event that can be measured and can take on different values indicating a quantity or a quality.

Appendices

Appendix 1 Race and Ethnicity as Variables

There are multiple ways in which race and ethnicity variables can be incorporated in study design or reporting. The table below summarizes some of the most common variable types and discusses if each type is appropriate for race and ethnicity when they are measured as a fixed identity descriptor.

Table 1: Incorporating categorical race and ethnicity variables into research design

VARIABLE TYPE	PURPOSE OF USE	APPROPRIATENESS*	EXAMPLE
Independent or grouping variable	To describe or compare groups, explain or predict outcomes.	Appropriate	A study describes graduation rates across different racial groups.
Moderator	To look at how race or ethnicity influences the direction or strength of relationship between other variables (i.e., interacts with other variables).	Appropriate	A study explores the impact of classroom size on graduation rates. Race is used as a moderating variable to see if the impact differs across racial groups.
Mediator	To see if race or ethnicity explains relationship between other variables.	Appropriate	A study explores the link between education and income. It uses race as a proxy for exposure to racial discrimination to explain why education is associated with higher income in one region but not in another region.
Control variable	To account for (remove) the effects of race or ethnicity on the relationship between other variables.	Appropriate	A study is interested in assessing the impacts of classroom size on graduation rates when the effects of race are accounted for.
Dependent variable	To see if other variables affect race or ethnicity.	NOT appropriate, unless they capture the fluidity of racial/ethnic identity over time. ²⁰	A study is interested in understanding how learning parental heritage language impacts the development of ethnic identity of immigrant children.

* In reference to the use of categorical race and ethnicity variables for the purpose listed.

Appendix 2: Case Studies

C1 Social nature of race and ethnicity variables.

Socially constructed concepts change their meaning and the way they are measured as social norms, values and beliefs change. The social nature of the concept of ethnicity is evident in changes to examples provided for the ethnic and cultural origin question in the Canadian census over the years.

Table 2: Ethnic and cultural origin categories in Canadian census by year

IN 1981, THE FOLLOWING 15 CATEGORIES WERE LISTED:

French	Italian	Chinese
English	Ukrainian	Inuit
Irish	Dutch (Netherlands)	Status or registered Indian
Scottish	Polish	Non-status Indian
German	Jewish	Métis

IN 2016, THE FOLLOWING 28 CATEGORIES WERE LISTED:

Canadian	Salish	Korean
English	Métis	Jamaican
Chinese	Inuit	Greek
French	Filipino	Iranian
East Indian	Irish	Lebanese
Italian	Dutch	Mexican
German	Ukrainian	Somali
Scottish	Polish	Colombian
Cree	Portuguese	
Mi'kmaq	Vietnamese	

IN 2021, THE ETHNIC AND CULTURAL ORIGIN QUESTION DID NOT LIST ANY EXAMPLES AND OFFERED A LIST OF 500 CATEGORIES AS A REFERENCE.

20 Categorical race and ethnicity variables can capture the fluidity of these identity factors when responses are tracked over time. This means a dataset would capture all responses ever recorded for the race or ethnicity question instead of tracking the most recent response.

C2 Justifying inclusion and exclusion of groups and using a medical standard for group comparisons.

Crawley and colleagues (2008) examined how perceived racial discrimination (as reported on a survey) affects colorectal and breast cancer screening rates among African American, American Indian/Alaskan Native, Asian and Latino populations in the U.S.

The authors justified the inclusion of race by citing reports that document evidence of race-based discrimination in the U.S. health-care settings. The authors also provided a reason for excluding White, non-Hispanic population from the study. This was done because there was no evidence documenting race-based discrimination against this group.

In their analytic and reporting technique, the authors compared cancer screening rates for each racial group to a medical standard - recommended intervals for colorectal and breast cancer screening.

The study found that perceived medical discrimination was associated with lower cancer screening rates among women and among men without a usual source of health care.

C3 Using race and ethnicity variables interchangeably.

Chiu and colleagues (2011) used Statistics Canada's National Population Health Survey (1996) and the Canadian Community Health Survey (2001, 2003, 2005) to identify "ethnic-specific BMI cut-off values for assessing diabetes risk" among the following "ethnic groups": White, Black, South Asian and Chinese.

Using these categories in reference to ethnic groups is problematic because they were not used in the original surveys to measure ethnicity. In fact, the National Population Health Survey questionnaire lists these categories as response options to a question asking to describe one's race or colour, while the Canadian Community Health Survey lists these response options to a question about one's cultural or racial background.

This study is an example of using a race variable as a substitute for ethnicity, which we do not advise. It also highlights problems with the data currently available on race and ethnicity. The national surveys used in this study do not measure one distinct concept and use a combination of racial, cultural and ethnic categories as response options.

C4 Inaccuracy of genetic attribution of race effects.

For a long-time, sickle cell disease (SCD) was considered a “Black disease” (Mason, 1922). Researchers today no longer attribute SCD to innate racial differences. Instead, they attribute it to differences in the geographic origin of one’s ancestors. It is understood that SCD is caused by a mutation in the beta-globin gene. This mutation was most common in geographic regions where malaria was prevalent, as the mutated gene gave carriers a survival advantage against malaria. These regions include Central and West Africa, the Middle East, and India. SCD is most common among populations descending from these parts of the world, regardless of their race (El-Hazmi et al., 2011). This means that being a certain race does not make one more susceptible to SCD, but having ancestors that descended from regions where malaria was prevalent does.

C5 Social attribution of race effects and discussion of mediators.

Wrigley-Field and colleagues (2021) investigated the relationship between racial and ethnic disparities in mortality and neighbourhood disadvantage in Minnesota before and during the COVID-19 pandemic.

Neighbourhood disadvantage was measured by the Area Deprivation Index, which includes indicators of income, education, employment and housing quality. While the study does not align with some recommendations presented in this guide (e.g., multiple racial and ethnic groups are homogenized into a Black, Indigenous and People of Colour [BIPOC] category), it offers a good example of social attribution of race and ethnicity effects through a strong discussion of contributing factors. One of the study’s key findings was the reversal of racial and ethnic disparity in total mortality as well as a change in its key drivers. Before the pandemic, the White population had higher mortality than the BIPOC population living in similarly disadvantaged neighbourhoods. This was largely explained by the neighbourhood selection effect, whereby White residents living in highly disadvantaged areas were more likely to be disadvantaged on other characteristics (e.g., having a criminal history) than their BIPOC counterparts.

During the pandemic, mortality rates for BIPOC residents increased by 41% and exceeded mortality rates for White residents. The authors attribute this reversal to structural inequities that made BIPOC residents more susceptible to COVID-19. These inequities include disproportionate exposure to occupational risks, exposure to denser housing, worse treatment by some medical providers and a higher likelihood of living in multigenerational households.

C6 Theoretical incorporation of mediators and moderators.

Neblett and colleagues (2010) propose a theoretical model explaining the increased risk of substance use initiation and substance use disorders among African American youth in the U.S. The model includes several mediators that explain racial disparities in substance use. Among these are exposure to race-based discrimination and the resulting stress that leads to unhealthy coping behaviors involving substance use.

The authors also propose a range of protective factors that can moderate the effects of stress on substance use. One of these factors is racial socialization, defined as conversations about race between parents and their children that aim to raise awareness of racism and discrimination and build a positive sense of racial identity.

C7 Simpson's paradox as a common error resulting from omission of moderators.

Simpson's paradox (also known as Yule-Simpson effect) commonly occurs when aggregated data is used to compare groups on several dimensions. The effect is present when the addition of moderator variables reverses or changes a relationship that appeared to be present at a higher level of aggregation.

For example, consider U.S.-reported COVID-19 case fatality rate (CFR) by race and Hispanic origin described in Table 3 below. The data is based on weekly reported cases and deaths from March 2020 to January 2023 and includes adults of all genders.

When the data is broken down by race and Hispanic origin only, it appears that the White, Non-Hispanic population had a higher case fatality rate than the Hispanic population (1.77% versus 1.1%). If further analysis is not done, one may conclude that the White, non-Hispanic population had a higher risk of death due to COVID-19 and proceed with identifying program and policy solutions.

Table 3: U.S.-reported COVID-19 case fatality rate (CFR) by race and Hispanic origin

RACE, HISPANIC ORIGIN	CFR
White, Non-Hispanic	1.77%
Hispanic	1.1%

However, the relationship between race, Hispanic origin, and case fatality rates reverses when data is further disaggregated by age. As demonstrated in Table 4 below, case fatality rates for the Hispanic population were, in fact, higher than for the White, non-Hispanic population at every age category.

In this case, age moderates the relationship between race, Hispanic origin and case fatality rates because it is one of the most significant risk factors for death due to COVID-19. And since the White, Non-Hispanic population in the U.S. is older than the Hispanic population, it is important to include it.

Table 4: U.S.-reported COVID-19 case fatality rate (CFR) by race and Hispanic origin disaggregated by age

AGE	WHITE, NON-HISPANIC CFR	HISPANIC CFR
18 - 29	0.03%	0.05%
30 - 39	0.1%	0.2%
40 - 49	0.3%	0.5%
50 - 64	1.1%	1.4%
65 - 74	3.3%	4.4%
75+	10.4%	11.4%

C8 Importance of including moderators.

Including moderators can help develop a more nuanced understanding of the experiences of Indigenous women and men in the Canadian labour market. Park (2021) used 2016 census data to analyze rates of over-qualification in Indigenous men and women with post-secondary education.

The author found that Indigenous men and women benefit differently from post-secondary education and experience different impacts of over-qualification. In general, over-qualified Indigenous women had lower income levels compared to over-qualified Indigenous men. This means that the negative financial impact of over-qualification was stronger for Indigenous women. At the same time, university level education seemed to be more advantageous to Indigenous women than to Indigenous men. This is because rates of over-qualification among Indigenous women with university-level education were lower than among Indigenous men.

C9 The complexities of comparing race and ethnicity variables from different datasets.

While the U.S. and Canadian censuses use several similar labels in their race-based variables (e.g., White and Black), there are also considerable differences in the way these variables are measured and processed.

There is a race variable in the U.S. census but not in the Canadian census which includes a visible minority variable instead. These two variables are not directly comparable. For example, while American Indian and Alaska Native populations are included in the U.S. race variable, Indigenous Peoples are not captured in the Canadian visible minority variable.

Another difference that adds complexity is how Hispanic populations are captured in the two countries.

In the U.S. census, Hispanic origin is captured in the ethnicity variable that can be grouped with any other race category (e.g., White Hispanic, Black Hispanic) in data processing.

In the Canadian census, Hispanic origin is captured in the visible minority variable. When a person selects 'Latin American' and another category (e.g., White or Black), the data is not processed in a way that keeps all selected categories visible by creating a new combined category like in the U.S. census (e.g., White Latin American or Black Latin American). Instead, if a person in Canada identifies as Latin American and White, they will be counted as White (non-visible minority). However, if they identify as Latin American and Black, they will be counted under 'multiple visible minorities' category that also includes some other (but not all) multiracial individuals.

These differences complicate direct comparisons of Hispanic, multiracial, and other populations using race-based census variables of the two countries and illustrate the importance of understanding how racial and ethnic data is collected and processed before using it.

C10 Data processing approaches for multiple response questions.

Table 5 below describes two common grouping strategies used to process responses of individuals who identify as multiracial by selecting more than one response category. This is usually done to avoid double counting responses received from the same person. Both approaches have limitations and work better for different types of projects.

Table 5: Grouping strategies used to process responses of individuals who identify as multiracial by selecting more than one response category

APPROACH	HOW IT WORKS	LIMITATIONS	WORKS BEST FOR
Creating a new “multiracial” category	All individuals who selected multiple response options are counted together.	Does not account for variability between multiracial individuals.	Projects interested in the experience of being multiracial (regardless of the specific race combinations).
Assigning each response to a single race category using bridging methods	Individuals who chose multiple response options are reassigned to existing categories using a set of rules (e.g., Black & White will be grouped with Black).	Does not account for variability between monoracial and multiracial individuals.	Projects interested in the experience of people who identify as a specific race (regardless of other identity dimensions).

C11 The importance of considering diversity within racial and ethnic groups.

Schwabish and Feng (2021) illustrate the dangers of grouping diverse ethnic groups by highlighting dramatic differences in poverty rates between the American Indian and Alaska Native populations in the U.S. For example, the Aleut Indigenous community has a poverty rate of 5.9%, whereas those who identify as part of the Sioux Native American tribe experience poverty rates of 36.9%. The authors explain that grouping these populations together would mask differences in poverty rates and ultimately impact how resources are distributed and how the groups are perceived in society.

C12 Examples of available guidelines on describing racial and ethnic groups.

Consider the following examples of guidelines on how to describe racial and ethnic groups (in chronological order).

Kaplan and Bennett (2003) discourage using depersonalizing plural nouns. For example, instead of “Asians”, they offer “Asian Americans” or “Asian/Pacific Islander Americans”. They further encourage using even more precise and descriptive terms such as “Cambodian-born U.S. residents,” “enrolled members of the Rosebud Sioux Tribe,” or “adults in the study sample who self-identified as White.” Finally, they suggest using quotation marks to set off the words “race” and “racial” to emphasize the imprecision of these terms.

Queen’s University (2019) published guidance on Indigenous terminology. “Indigenous”, “First Nations”, “Métis” and “Inuit” are listed as preferred terminology and terms to avoid include “Native”, “Our Native People”, “Canada’s Native People”, “Indigenous Canadian”, “Aboriginal” and “Indian”.

The American Medical Association (2020) recommends avoiding terms like “other” or “non-Whites” as they are non-specific and imply White normativity. In the event the term is used because it has been prespecified, they advise to list all groups within the category. In its updated guidance (Flanagin et al., 2020) the Association discourages the use of “minority” as a stand-alone noun and suggests adding modifiers (e.g., “ethnic minority individuals”) or using “underserved”, “underrepresented” or “minoritized” where appropriate.

The Government of British Columbia (2021) published a [Writing Guide for Indigenous Content](#) that provides a range of recommendations and resources for writers. For example, the guide recommends using terminology preferred by the Indigenous people involved in the project and identifying individuals by their Nation instead of using the generic “Indigenous” term. It also cautions against using passive language when writing about Indigenous people and recommends using active language instead.

The Canadian Medical Association Journal (Stanbrook & Salami, 2023) recommends capitalizing names of racial, tribal and ethnic groups and listing racial and ethnic categories based on empiric rationale, rather than automatically listing “White” first.

C13 Illustrating within-group variability through data visualization.

The figures below use simulated individual-level data to display annual employment income for White, Latin American, Chinese and Black populations in B.C. The simulation is based on income averages reported in Statistics Canada's 2021 census tables. Income distribution depicted in the figures does not reflect reality because it is based on a set of assumptions that were made to generate the data. For this reason, individual incomes do not approach \$0 and do not exceed \$150,000.

Jitter plots (Fig. 1) help visualize within-group variability by showing where each person stands in relation to the group average. For privacy purposes, each dot can be made to represent small groups of people. In comparison, bar charts (Fig. 2) make it seem like people within the group fall behind their group mean and mask significant overlaps in income between individuals belonging to different groups.

In this example, jitter plot helps illustrate that there are significant differences in income within each population group as well as similarities between people belonging to different population groups.

Fig 1. 2020 Average annual employment income in British Columbia.

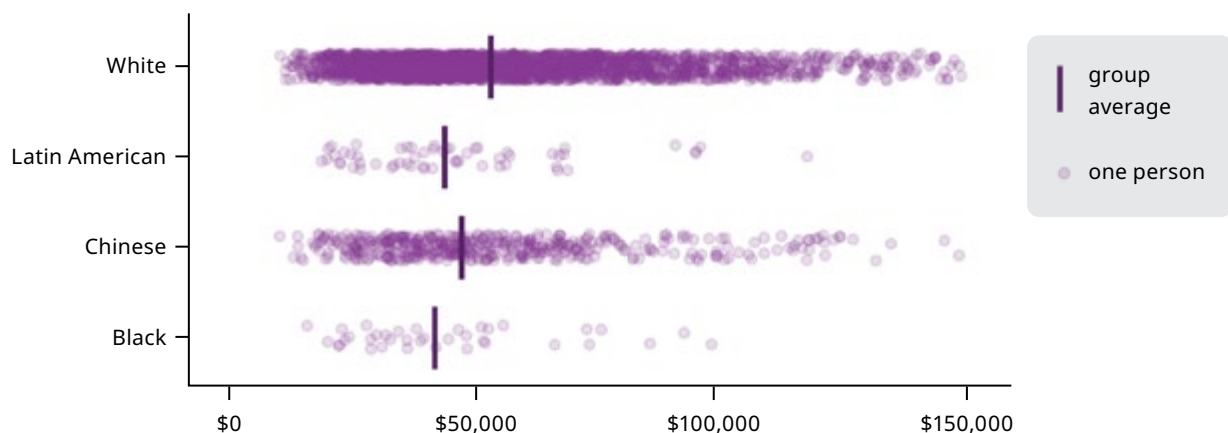


Fig 2. 2020 Average annual employment income in British Columbia.



C14 Discussing limitations to generalizability of findings that use census benchmarking.

Census benchmarking is frequently used in studies of racial and ethnic disparities, such as those examining over-representation of certain racial groups in interactions with the police. Applied to this field, this analytic technique compares the demographic profile of individuals interacting with the police to the demographic profile of residents in that police jurisdiction as determined by census data.

Census benchmarking has lower accuracy compared to other forms of benchmarking used in the field.²⁰ This is because, among other things, census data is often not available at the right level of granularity, does not account for mobility throughout the day and includes all residents regardless of their risk of a police encounter.

In a recent examination of racial disparities in police statistics from British Columbia, Wortley (2021) used census benchmarking to describe various patterns of racial over-representation in citizen-police interactions. Throughout the report, the author made it clear that the findings apply only to the police jurisdictions that were examined.

The author also recognized limitations of the benchmarking method used by stating that its accuracy could not be determined. He further explained that the accuracy of racial disparities revealed in the study could have been impacted by individuals who reside in jurisdictions that were out of scope of the study (and, therefore, would not be in the census counts used) but interacted with the police agencies that were in scope (and, therefore, would be in the police statistics used).

²⁰ Fridell (2004).

Appendix 3: Self-Assessment Checklist

This checklist can help data users assess the level to which relevant recommendations discussed in the guide have been considered and implemented in their projects.

The questions are worded from a first-person perspective for relatability. However, this does not mean that all decisions should be made by one person. Multiple individuals associated with the project are usually involved in the decision-making process.

The questions pertain only to the recommendations provided in the guide and do not cover other important considerations required in projects about racial and ethnic groups (e.g., community benefit and involvement, Indigenous data sovereignty rights and personal qualifications and positionality).

R1 Define race and ethnicity as distinct social concepts and provide a justification for their use.

R1.1. How am I defining race or ethnicity?

R1.2. Do I have a justification for using race or ethnicity in this project?

R1.3. Am I using a conceptually different variable as a substitute for race and ethnicity?

R1.4. Am I using race or ethnicity as a substitute for a conceptually different measurable variable?

R1.5. Is there a related concept that is better suited for this project?

R1.6. Where can I document the concept definition and justification for my audience?

R2 Do not attribute the effects of race and ethnicity to biological or genetic variation in racial and ethnic groups.

R2.1. Am I using race or ethnicity to make conclusions about innate biological or genetic differences between racial or ethnic groups?

R2.2. Am I describing the project or its findings in a way that may lead my audience to make genetic attributions of race or ethnicity effects?

R3 Consider conceptually relevant individual, social, structural and contextual factors that explain and modify associations between race, ethnicity and other variables. Explicitly identify what factors are included and excluded.

R3.1. Have I considered factors that explain why there may be a relationship between race or ethnicity and an outcome variable I am interested in (mediators)?

R3.2. Have I considered factors that may change the relationship between race or ethnicity and an outcome variable I am interested in (moderators)?

R3.3. Can I include any mediators or moderators in my data analysis?

R3.4. Can I provide more information about potential mediators or moderators that I am not including in the analysis?

R3.5. Where can I document information about mediators and moderators that I am including and that I am not including?

R4 Consider conceptually relevant individual, social, structural and contextual factors that explain and modify associations between race, ethnicity and other variables. Explicitly identify what factors are included and excluded.

R4.1. How were race or ethnicity measured in the dataset(s) I am thinking of using?

R4.2. Does the way race or ethnicity were measured the dataset(s) align with my concept definitions?

R4.3. What aspects of race or ethnicity were not measured in the dataset(s) and how may this impact the results?

R4.4. Where can I document information about the data collection method(s) and its implications for the results?

R5 If grouping is necessary, ensure that resulting categorizations meaningfully represent racial and ethnic groups included.

R5.1. What was the underlying principle I used to create new groupings?

R5.2. Are the new groupings meaningful for the purpose of this project?

R5.3. Were the new groups created from different datasets? Were the categories in these datasets comparable and fit to be combined?

R5.4. Do the new groupings result in overly broad categories that combine groups that have little in common and are hard to interpret?

R5.5. Where can I document the grouping decisions made in the project?

R6 When describing racial and ethnic groups, use terminology preferred by their members and appropriate in the context of the project.

R6.1. Am I using appropriate terminology to describe new groupings and in my data tables, visualizations, and/or report(s)?

R6.2. Did I make any changes to the original labels used in the dataset? Have I considered the risks and benefits of changing labels?

R6.3. Where can I document labeling decisions made in this project?

R7 When some racial and ethnic groups are included and others are excluded, explain why this is done.

R7.1. Do I have a justification for including some groups and excluding other groups in the analysis?

R7.2. Am I excluding some groups only because of a small sample size? Can I take any steps to include them in this project or in the future?

R7.3. Where can I document the justification for including some groups and excluding others?

R8 When comparing racial and ethnic groups, explain why this is done and take steps to mitigate the risk of harm from drawn comparisons.

R8.1. Do I have a justification for comparing racial or ethnic groups?

R8.2. Did I include moderators to check if differences or lack of differences between groups remain?

R8.3. How am I interpreting the presence or absence of differences?

R8.4. What referent group or value am I using for comparison and why?

R8.5. How can I help my audience interpret the findings correctly and not perpetuate stereotypes?

R8.6. Have I considered the risks and benefits of publicly releasing the findings?

R8.7. Where can I document the justification for comparison, the reason why a certain referent group or value was chosen and the appropriate interpretation of findings?

R9 Include information explaining race and ethnicity-related results, their accuracy and whether they can be generalized.

R9.1. How am I interpreting race or ethnicity-related findings?

R9.2. Are included and excluded moderating and mediating factors considered in my interpretation?

R9.3. To what specific groups, places and time can the findings be generalized?

R9.4. What are the limitations of the data and analytic methods used that may impact the accuracy of the findings?

R9.5. Where can I document my interpretation of the findings and project limitations that my audience should be aware of?



References

References

- Agadjanian, A. (2022). How many Americans change their racial identification over time? *Socius*, 8.
- AMA Manual of Style Committee. (2020). *AMA manual of style: A guide for authors and editors*.
- American Psychological Association. (2019). Race and Ethnicity Guidelines in Psychology: Promoting Responsiveness and Equity.
- BC Office of the Human Rights Commissioner. (2020). Disaggregated demographic data collection in British Columbia: The grandmother perspective. Retrieved from <https://bchumanrights.ca/wp-content/uploads/BCOHR Sept2020 Disaggregated-Data-Report FINAL.pdf>
- Brasel, K. J., & Coimbra, R. (2022). Toward improved reporting and use of race, ethnicity, sex, and gender: An editorial. *Journal of Trauma and Acute Care Surgery*, 92(4), 629-630.
- Braveman, P., Cubbin, C., Marchi, K., Egerter, S., & Chavez, G. (2001). Measuring socioeconomic status/position in studies of racial/ethnic disparities: maternal and infant health. *Public Health Reports*, 116(5), 449.
- British Columbia (2021). Writing Guide for Indigenous Content. Retrieved from <https://www2.gov.bc.ca/gov/content/governments/services-for-government/service-experience-digital-delivery/web-content-development-guides/web-style-guide/writing-guide-for-indigenous-content>
- Brown, T. N., Donato, K. M., Laske, M. T., & Duncan, E. M. (2013). Race, nativity, ethnicity, and cultural influences in the sociology of mental health. *In Handbook of the Sociology of Mental Health* (pp. 255-276). Springer, Dordrecht.
- Campbell, M. E., & Troyer, L. (2007). The implications of racial misclassification by observers. *American Sociological Review*, 72(5), 750-765.
- Chiu, M., Austin, P. C., Manuel, D. G., Shah, B. R., & Tu, J. V. (2011). Deriving ethnic-specific BMI cutoff points for assessing diabetes risk. *Diabetes Care*, 34(8), 1741-1748.
- Clarke, D. E., Colantonio, A., Rhodes, A. E., & Escobar, M. (2008). Ethnicity and mental health: Conceptualization, definition and operationalization of ethnicity from a Canadian context. *Chronic Diseases in Canada*, 28(4), 128-147.
- Cokley, K., & Awad, G. (2007). Conceptual and methodological issues related to multicultural research. In P. Heppner, D. Kivlighan, & B. Wampold (Eds.), *Research Design in Counseling* (3rd ed., pp. 366-384). Belmont, CA: Brooks Cole.
- Cokley, K., & Awad, G. H. (2013). In defense of quantitative methods: using the "master's tools" to promote social justice. *Journal for Social Action in Counseling & Psychology*, 5(2), 26-41.
- Comstock, R. D., Castillo, E. M., & Lindsay, S. P. (2004). Four-year review of the use of race and ethnicity in epidemiologic and public health research. *American Journal of Epidemiology*, 159(6), 611-619.
- Cooper, R., & David, R. (1986). The biological concept of race and its application to public health and epidemiology. *Journal of Health Politics, Policy and Law*, 11(1), 97-116.

- Crawley, L. M., Ahn, D. K., & Winkleby, M. A. (2008). Perceived medical discrimination and cancer screening behaviors of racial and ethnic minority adults. *Cancer Epidemiology Biomarkers & Prevention, 17*(8), 1937-1944.
- Datta, G., Siddiqi, A., & Lofters, A. (2021). Transforming race-based health research in Canada. *CMAJ, 193*(3), E99-E100.
- Davis, J. E. (1992). Reconsidering the use of race as an explanatory variable in program evaluation. *New Directions for Program Evaluation, 1992*(53), 55-67.
- Davis, L. P., & Museus, S. D. (2019). What is deficit thinking? An analysis of conceptualizations of deficit thinking and implications for scholarly research. *NCID Currents, 1*(1).
- Dirette, D. P. (2014). Questions about race as a research variable. *The Open Journal of Occupational Therapy, 2*(3), 1.
- El-Hazmi, M. A., Al-Hazmi, A. M., & Warsy, A. S. (2011). Sickle cell disease in Middle East Arab countries. *Indian Journal of Medical Research, 134*(5), 597-610.
- Feliciano, C. (2016). Shades of race: How phenotype and observer characteristics shape racial classification. *American Behavioral Scientist, 60*(4), 390-419.
- Flanagin, A., Frey, T., Christiansen, S. L., & AMA Manual of Style Committee. (2021). Updated guidance on the reporting of race and ethnicity in medical and science journals. *Jama, 326*(7), 621-627.
- Fridell, L. A. (2004). *By the numbers: A guide for analyzing race data from vehicle stops*. Washington, DC: Police Executive Research Forum.
- Fullilove, M. T. (1998). Comment: abandoning "race" as a variable in public health research--an idea whose time has come. *American Journal of Public Health, 88*(9), 1297-1298.
- Garcia, D., & Abascal, M. (2016). Colored perceptions: racially distinctive names and assessments of skin color. *American Behavioral Scientist, 60*(4), 420-441.
- Garcia, J. A. (2017). The race project: researching race in the social sciences researchers, measures, and scope of studies. *Journal of Race, Ethnicity, and Politics, 2*(2), 300-346.
- Harell, A., & Lieberman, E. (2021). How information about race-based health disparities affects policy preferences: evidence from a survey experiment about the COVID-19 pandemic in the United States. *Social Science & Medicine, 277*, 113884.
- Holder, E., & Xiong, C. (2022). Dispersion vs disparity: hiding variability can encourage stereotyping when visualizing social outcomes. *IEEE Transactions on Visualization and Computer Graphics*.
- Ioannidis, J. P., Powe, N. R., & Yancy, C. (2021). Recalibrating the use of race in medical research. *Jama, 325*(7), 623-624.
- James, A. (2001). Making sense of race and racial classification. *Race and Society, 4*(2), 235-247.
- Jugert, P., Kaiser, M. J., Ialuna, F., & Civitillo, S. (2022). Researching race-ethnicity in race-mute Europe. *Infant and Child Development, 31*(1), e2260.
- Kaplan J. & Bennett T. (2003). Use of race and ethnicity in biomedical publication. *Jama, 289*(20):2709-16.

- Kaufman, J. S., & Cooper, R. S. (2001). Commentary: considerations for use of racial/ethnic classification in etiologic research. *American Journal of Epidemiology*, 154(4), 291–298.
- Kelly, Y., Panico, L., Bartley, M., Marmot, M., Nazroo, J., & Sacker, A. (2009). Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. *Journal of Public Health*, 31(1), 131-137.
- Khan, A. T., Gogarten, S. M., McHugh, C. P., Stilp, A. M., Sofer, T., Bowers, M. L., ... & Nelson, S. C. (2022). Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. *Cell Genomics*, 100155.
- Krieger N. (2003). Does racism harm health? Did child abuse exist before 1962? On explicit questions, critical science, and current controversies: an ecosocial perspective. *American Journal of Public Health*, 93, 194-199.
- Krieger, N. (2005). Stormy weather: race, gene expression, and the science of health disparities. *American Journal of Public Health*, 95(12), 2155-2160.
- Krieger, N., Rowley, D. L., Herman, A. A., Avery, B., & Phillips, M. T. (1993). Racism, sexism, and social class: implications for studies of health, disease, and well-being. *American Journal of Preventive Medicine*, 9(6), 82-122
- LaVeist, T. A. (1994). Beyond dummy variables and sample selection: what health services researchers ought to know about race as a variable. *Health Services Research*, 29(1), 1.
- LaVeist, T. A. (1996). Why we should continue to study race... but do a better job: an essay on race, racism and health. *Ethnicity & disease*, 6(1/2), 21-29.
- Lett, E., Asabor, E., Beltrán, S., Cannon, A. M., & Arah, O. A. (2022). Conceptualizing, contextualizing, and operationalizing race in quantitative health sciences research. *The Annals of Family Medicine*, 20(2), 157-163.
- Liebler C. A., Halpern-Manners A. (2008). A practical approach to using multiple-race response data: a bridging method for public-use microdata. *Demography*. 45(1):143-55.
- Lindeborg, M., Din, T., Araya-Quezada, C., Lawal, S., Heer, B., Rajaguru, P., ... & Fagan, J. (2022). Race and ethnicity in otolaryngology academic publications. *Otolaryngology - Head and Neck Surgery*, 166(6), 1196-1203.
- Link B. G., Phelan J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*. 38(suppl):80–94
- Jones, C. (2002). The impact of racism on health. *Ethnicity & Disease*, 12(1), 10-13.
- Mason, V. R. (1922). Sickle cell anemia. *Journal of the American Medical Association*, 79(16), 1318-1320.
- McKenzie, K. J., & Crowcroft, N. S. (1994). Race, ethnicity, culture, and science. *BMJ*, 309(6950), 286-287.
- McLemore, M. R. (2021). Reimagining methodological considerations for research studies using 'big' administrative data sets. *Paediatric and Perinatal Epidemiology*, 35(4), 491-492.
- Mickelson, R. A. (2003). When are racial disparities in education the result of racial discrimination? A social science perspective. *Teachers College Record*, 105(6), 1052-1086.

- Mohottige, D., Boulware, L. E., Ford, C. L., Jones, C., & Norris, K. C. (2022). Use of race in kidney research and medicine: concepts, principles, and practice. *Clinical Journal of the American Society of Nephrology*, 17(2), 314-322.
- Morgenstern, H. (1997). Defining and explaining race effects. *Epidemiology*, 8(6), 609-611.
- Moscou, S. (2008). The conceptualization and operationalization of race and ethnicity by health services researchers. *Nursing Inquiry*, 15(2), 94-105.
- Mustillo, S. A., Krieger, N., Gunderson, E. P., Sidney, S., McCreath, H., & Kiefe, C. I. (2004). The association of self-reported experiences of racial discrimination with black/white differences in preterm delivery and low birth weight: the CARDIA study. *American Journal Public Health*, 94, 2125-2131.
- Neblett Jr, E. W., Terzian, M., & Harriott, V. (2010). From racial discrimination to substance use: the buffering effects of racial socialization. *Child Development Perspectives*, 4(2), 131-137.
- Nelson, A. H., Jenkins, D., Zanti, S., Katz, M., & Berkowitz, E. (2020). *A Toolkit for Centering Racial Equity throughout Data Integration*. Actionable Intelligence for Social Policy. University of Pennsylvania.
- Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., ... & Gee, G. (2015). Racism as a determinant of health: a systematic review and meta-analysis. *PloS one*, 10(9), e0138511.
- Park, J. (2021). Over-qualification in the workforce: do Indigenous women and men benefit equally from high levels of education? *Aboriginal Policy Studies*, 9(2).
- Phinney, J. S. (1992). The multigroup ethnic identity measure: a new scale for use with diverse groups. *Journal of Adolescent Research*, 7(2), 156-176.
- Queen's University. (2019). Indigenous Terminology Guide. Retrieved from <https://www.queensu.ca/indigenous/ways-knowing/terminology-guide>
- Root, M. (2003). The Use of Race in Medicine as a Proxy for Genetic Differences. *Philosophy of Science*, 70(5), 1173-1183. doi:10.1086/377398
- Ross, P.T., Hart-Johnson, T., Santen, S.A. et al. (2020). Considerations for using race and ethnicity as quantitative variables in medical education research. *Perspectives in Medical Education*. 9, 318-323.
- Schwabish J. & Feng, A. (2021). Combining racial groups in data analysis can mask important differences in communities. Retrieved from <https://www.urban.org/urban-wire/combining-racial-groups-data-analysis-can-mask-important-differences-communities>
- Skinner-Dorkenoo, A. L., Sarmal, A., Rogbeer, K. G., André, C. J., Patel, B., & Cha, L. (2022). Highlighting COVID-19 racial disparities can reduce support for safety precautions among White US residents. *Social Science & Medicine*, 301, 114951.
- Social and Housing Statistics Section, Demographic and Social Statistics Branch, United Nations Statistics Division. (2003). *Ethnicity: A Review of Data Collection and Dissemination*.
- Stanbrook M. B. & Salami B. (2023). CMAJ's new guidance on the reporting of race and ethnicity in research articles. *CMAJ*. 195(6), E236-E238.

- Statistics Canada. (2019). Collecting Census Data on Canada's Visible Minority Population: A Historical Perspective. Reference period: 1981 to 1991. Retrieved from <https://publications.gc.ca/site/eng/9.815287/publication.html>
- Statistics Canada. (2017). Ethnic Origin Reference Guide, Census of Population, 2016. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2016/ref/guides/008/98-500-x2016008-eng.cfm>
- Statistics Canada. (2022). Ethnic or Cultural Origin Reference Guide, Census of Population, 2021. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2021/ref/98-500/008/98-500-x2021008-eng.cfm>
- Statistics Canada. (2022). Table 98-10-0439-01 Employment income statistics by visible minority, highest level of education, immigrant status and income year: Canada, provinces and territories, census metropolitan areas and census agglomerations with parts. Retrieved from <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810043901>
- Statistics Canada. (2022). Table 98-10-0324-01 Visible minority and population group by generation status: Canada, provinces and territories, census metropolitan areas and census agglomerations with parts. Retrieved from <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810032401>
- Statistics Canada. (2022). Table 11-10-0241-01 Low-income cut-offs (LICOs) before and after tax by community size and family size, in current dollars.
- Statistics Canada. (2022). Visible Minority and Population Group Reference Guide, Census of Population, 2021. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2021/ref/98-500/006/98-500-x2021006-eng.cfm>
- Stephens-Dougan, L. (2022). White Americans' Reactions to Racial Disparities in COVID-19. *American Political Science Review*, 1-8.
- Susarla, H. K., Dentino, K. M., Kalenderian, E., & Ramoni, R. B. (2014). The reporting of race and ethnicity information in the dental public health literature. *Journal of Public Health Dentistry*, 74(1), 21-27.
- United States Census Bureau. (2020). Measuring Racial and Ethnic Diversity for the 2020 Census. Retrieved from <https://www.census.gov/newsroom/blogs/random-samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html>
- Viano, S., & Baker, D. J. (2020). How administrative data collection and analysis can better reflect racial and ethnic identities. *Review of Research in Education*, 44(1), 301-331.
- Williams, D. R. (1994). The concept of race in health services research: 1966 to 1990. *Health Services Research*, 29(3), 261.
- Williams, D. R. (1996). Race/ethnicity and socioeconomic status: measurement and methodological issues. *International Journal of Health Services*, 26(3), 483-505.
- Williams, D. R. (1999). Race, socioeconomic status, and health. The added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, 896: 173-188.

- Wortley S., (2021). *Racial disparities in British Columbia police statistics: a preliminary examination of a complex issue*. Retrieved from https://bchumanrights.ca/wp-content/uploads/Wortley_Sep2021_Racial-disparities-police-statistics.pdf
- Wrigley-Field, E., Garcia, S., Leider, J. P., & Van Riper, D. (2021). COVID-19 Mortality at the neighborhood level: racial and ethnic inequalities deepened in Minnesota in 2020. *Health Affairs*, 40(10), 1644-1653.



Ministry of
Citizens' Services