

UNIVERSITY INSTITUTIONAL REVIEW BOARD

Collecting and Breaking Down Data by Demographic Variables: **Why and How?**



SACRAMENTO
STATE

Contents

- Purpose 4**
- Why Collect Demographic Information? 5**
- Importance of Breaking Down Data by Subgroups 5**
 - In education: 5**
 - In medicine and public health: 5**
- Concerns About Collecting Demographic Information 6**
- General Tips for Creating Inclusive Demographic Questions 6**
- Explanations of Some Demographic Terms..... 7**
 - Sex, Gender, Gender Identity, Cisgender/Transgender, Sexual Orientation 7**
- Sample Inclusive Questions..... 7**
 - Further reading on how to construct inclusive demographic questions 9**
 - Readings on the importance of disaggregating data 9**
- References 9**



Purpose

In the spirit of the Belmont Report's ethical principles of autonomy, beneficence, and justice, the IRB thought it prudent to remind and further edify researchers about the reasoning and potential inclusiveness of collecting certain demographic information. It is not comprehensive nor a stipulation that such demographics be collected in all research. This document is meant as a guide and suggestion for researchers to think about the reasoning for choosing their population, sampling, and question(s) of race, gender, class, etc. Researchers are encouraged to inquire beyond this guide for more specifics, given the demographics recommended for their study. One will also need to consider how to compare one's own findings with previous research that may have used different terminology and/or categories for demographic variables.

We recognize that this is a dynamic and changing document, and we invite all feedback to be sent to irb@csus.edu.





Why Collect Demographic Information?

1. To test research questions/ hypotheses
2. To describe their sample
 - a. To assess the representativeness of their sample to the population they were seeking to study
 - b. To facilitate a comparison between their findings and the findings of other studies
 - c. To allow others to test the replicability of their findings
 - d. To understand the limitations of the generalizability of their findings
3. To identify whether the findings differ by subgroups such as race, gender, class, etc.
 - a. This is primarily relevant for larger, quantitative studies.
 - b. The principle of beneficence (e.g., maximize benefits, and minimize harm) should guide the decision to analyze data by sub-group, with particular consideration to ways in which the data might be misused (e.g., early studies on racial differences in intelligence were used to bolster the eugenics movement).
 - c. "When describing differences between groups of people, focus on the qualities that are relevant to the situation at hand. For example, in a study of sex chromosome-linked illnesses, study participants' biological sexes are probably relevant, while participants' sexual orientations are probably not." (APA, 7th ed.)

Importance of Breaking Down Data by Subgroups

When certain populations of people are included but responses are not disaggregated by identity, we are unable to discern whether the research findings apply to them. These groups are thus left out of the benefits of the research. The dangers of failing to attend to important differences between groups has been illustrated across multiple fields (see "Be Data Literate Part I: How Aggregated Data Misleads Management," 2018 below for more examples):

In education:

"one district's state test data indicated that eighth-grade math scores steadily improved over three years. When the data team disaggregated those data, they discovered that boys' scores improved, while girls' scores actually declined. Another school noticed increased enrollment in their after-school science club. However, disaggregated data indicated that minority students, even those in more advanced classes, weren't signing up." (Mather, 2012)

In medicine and public health:

The persistently worse outcomes that women experience in the diagnosis and treatment of heart attacks (Alabas et al., 2017) can be traced, in part, back to the fact that, until recently, medical research was conducted exclusively with human (and animal) males and, even now, researchers often fail to disaggregate by gender, with significant negative implications on health outcomes for women (Runnels et al., 2014). Furthermore, approximately 20% of drugs recently tested for approval demonstrated different responses across racial/ethnic groups (Ramamoorthy et al., 2015), showing the importance of identifying our samples and disaggregating the data.



In each case, until researchers collected and disaggregated the data, the assumption was that the research findings applied to the whole sample and would, therefore, apply to all humans equally. This assumption was maintained because it was never tested, allowing inequities in health and educational outcomes to remain unnoticed. [It also worth cautioning against interpreting any group differences as biologically-based without the consideration of social, economic, and other environmental factors that often provide more compelling explanations.]

Concerns About Collecting Demographic Information

1. Potential loss of privacy on behalf of the participant (the more demographic data collected, the easier it is to identify the participant).
 - a. We can mitigate this concern by taking the appropriate precautions to ensure that only the researcher can access the data.
2. Identity is complex and asking participants to classify themselves into categories that do not match their identities can cause frustration as well as uncertainty about how to respond
 - a. “In general, respect the language that people use to refer to themselves, and understand that the language used to refer to certain groups of people can and does change over time. Recognize also that group members may not always express total agreement about this language.” (APA, 7th ed)
 - b. Thus, it is important to be intentional about how we ask these questions. See the tips and examples below.

General Tips for Creating Inclusive Demographic Questions

1. Given that forced choice responses are inherently limited and cannot cover the entire gamut of social identities, participants should be given the option to write in a response that best reflects their identity. These options are best phrased as “Another gender _____” or “Something else _____.” The use of the term “Other” is discouraged as it can be perceived as alienating.
2. Participants should be given the option of not responding to any (or all) of the demographic questions.
3. When multiple response options may be valid for a single participant, participants should be given the option to choose all that apply.



Explanations of Some Demographic Terms

Race and Ethnicity

Race and ethnicity are social constructs used to categorize groups of people. There exists little agreement about how to distinguish the two concepts. The names of racial and ethnic categories as well as the defining characteristics of each category vary across time and place. Generally speaking, race in the US is often based on geographic location of one's ancestors as well as phenotypic traits such as skin color. Ethnicity is an identity based in shared ancestry, culture, language, religion, beliefs, and/or customs. To learn more: <https://genderedinnovations.stanford.edu/terms/race.html>

Sex, Gender, Gender Identity, Cisgender/Transgender, Sexual Orientation

Sex is a way of classifying people (e.g., male, female, intersex) based on genetics, hormone levels, and physical traits, such as primary and secondary sexual characteristics. Sex is typically assigned at birth, based on observation of external genitalia, although it is significantly more complex and involves many factors. To learn more: <https://genderedinnovations.stanford.edu/terms/sex.html>

Gender refers to the social and cultural characteristics associated with women, men, boys and girls, such as behaviors and appearances.

Gender identity is one's internal experience of one's gender, which may coincide with or differ from one's sex assigned at birth. To learn more: <https://genderspectrum.org/articles/understanding-gender>

Cisgender refers to individuals whose gender identity corresponds to their assigned sex.

Transgender refers to individuals whose gender identity differs from their assigned sex. Gender identity and transgender status should be assessed separately.

Sexual orientation is an interpersonal identity that describes who we are physically, emotionally and/or romantically attracted to (or not). Although many labels exist for these categories, some of the more commonly used labels include gay, lesbian, bisexual, pansexual, asexual, queer, and heterosexual.





Sample Inclusive Questions

Please note that these are suggestions and can be modified to best fit the context of your study. The choice of which questions to include may be guided by theory and previous research.

1. Which best describes your gender? Choose all that apply.
 - Woman
 - Man
 - Non-binary
 - Something else _____
 - Prefer not to answer
2. Do you identify as transgender?
 - Yes
 - No
 - Unsure _____
 - Prefer not to answer
3. Which of the following best represents your racial/ ethnic heritage? Choose all that apply.
 - Asian
 - African/Black
 - Hispanic/Latinx
 - Middle Eastern/ North African
 - Native American
 - Native Hawaiian / Pacific Islander
 - White / European
 - Something else _____
 - Prefer not to answer
4. Which best describes your sexual orientation? Choose all that apply.
 - Asexual
 - Bisexual
 - Gay/Lesbian
 - Heterosexual
 - Pansexual
 - Queer
 - Sexually fluid
 - Something else _____
 - Prefer not to answer



5. Age_____ or prefer not to answer [Asking for a specific age rather than a DOB is preferred because it provides less identifying information e.g., 45 years old versus 5/4/78. However, if your sample size is small, you may want to consider using categories instead e.g., 18-25 years old, 26-30 years old, etc.]
6. What is the highest degree or level of school you have completed? (If you're currently enrolled in school, please indicate the highest degree you have received.)
 - Some high school
 - High school degree or equivalent (e.g. GED)
 - Some college
 - Associate degree (e.g. AA, AS)
 - Bachelor's degree (e.g. BA, BS)
 - Master's degree (e.g. MA, MS, MEd)
 - Professional degree (e.g. MD, DDS, DVM)
 - Doctorate (e.g. PhD, EdD)
 - Prefer not to answer
7. Other demographic questions that may be useful to ask (this is not an exhaustive list; see Hughes, Camden, & Yangchen, 2016, cited below, for examples):
 - Social class
 - Religion
 - Relationship status
 - Employment status
 - Parental status / caretaking status
 - Disability
 - Location

Further reading on how to construct inclusive demographic questions

1. Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., Kirn, A., Benson, L., & Potvin, G. (2016). More Comprehensive and Inclusive Approaches to Demographic Data Collection. *School of Engineering Education Graduate Student Series. Paper 60*. Retrieved from <http://docs.lib.purdue.edu/enegs/60>
2. Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research*, 21(3), 138–151. Retrieved from <https://search-ebSCOhost-com.proxy.lib.csus.edu/login.aspx?direct=true&db=psyh&AN=2016-45623-001>
3. Rosenberg, S. (2017, Mar 13). Respectful Collection of Demographic Data. Retrieved from <https://medium.com/@anna.sarai.rosenberg/respectful-collection-of-demographic-data-56de9fcb80e2>



Readings on the importance of disaggregating data

1. Mather, M. A. (2012, March 15). Using Data Tip #9: Disaggregating Data Makes the Invisible Visible. Retrieved from <https://usingdata.wordpress.com/2012/03/15/using-data-tip-9-disaggregating-data-makes-the-invisible-visible/>
2. Sharpe, R.V. (2019). Disaggregating data by race allows for more accurate research. *Nature Human Behavior*, 3, 1240. doi:10.1038/s41562-019-0696-1
3. Be Data Literate Part I: How Aggregated Data Misleads Management (2018, Sept 5). Retrieved from <https://www.customercontactweekdigital.com/drucker/decision-making/articles/be-data-literate-part-i-how-aggregated-data-misleads-management>

References

1. Alabas, O. A., Gale, C. P., Hall, M., Rutherford, M. J., Szummer, K., Lawesson, S. S., Jernberg, T. (2017). Sex Differences in Treatments, Relative Survival, and Excess Mortality Following Acute Myocardial Infarction: National Cohort Study Using the SWEDEHEART Registry. *Journal of the American Heart Association*, 6(12). doi:10.1161/jaha.117.007123
2. American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.
3. Barthelme, S (2009, Feb 9). Culture and Perception, part II: The Muller-Lyer illusion. Retrieved from <http://cognitionandculture.net/blogs/simons-blog/culture-and-perception-part-ii-the-muller-lyer-illusion/>
4. Ramamoorthy, A., Pacanowski, M., Bull, J. and Zhang, L. (2015). Racial/ethnic differences in drug disposition and response: Review of recently approved drugs. *Clinical Pharmacology and Therapeutics*, 97, 263-273. doi:[10.1002/cpt.61](https://doi.org/10.1002/cpt.61)
5. Runnels, V., Tudiver, S., Doull, M., & Boscoe, M. (2014). The challenges of including sex/gender analysis in systematic reviews: a qualitative survey. *Systematic reviews*, 3, 33. doi:10.1186/2046-4053-3-33

Much of the 'Why Collect Demographic Information?', 'Concerns About Collecting Demographic Information,' and the 'Sample Inclusive Questions' sections are taken and modified from Hughes, J. L., Camden, A. A., & Yangchen, T. (2016).

Written by Sharon M. Flicker, Matias Castro, and Zach C. Schudson

Acknowledgment: Leah Vargas