Understanding the Strengths & Limitations of Secondary Data in Gender/Sex Epidemiological Analysis

Developed in Women, Gender and Health 207:
Advanced Topics of Women, Gender, and Health,
Harvard School of Public Health, Spring 2024

Course Instructed by Sabra L. Katz-Wise and Kieran Todd

Teaching Example Authored by
Emily Newton-Hoe, Meekang Sung, & Seetha Davis

Appropriate HSPH Core Courses:
This teaching example is meant to be used in introductory biostatistics and epidemiology courses, such as **EPI 202 – Epidemiologic Methods 2: Elements of Epidemiologic Research** that teach exploratory data analysis using publicly available datasets.

Students' Prior Knowledge:
Introductory biostatistics and epidemiology courses like EPI 202 familiarize students with core epidemiologic concepts such as confounding, interaction, and effect measure modification. These courses also provide students with opportunities to practice analyzing publicly available datasets. Thus, by the time students engage with this teaching example, they should be prepared to effectively apply these concepts in real-world scenarios.

Brief Background:
A prevalent issue in gender/sex research is the lack of clarity regarding the specific aspects of gender relations and sex-linked biology that are pertinent to health outcomes. By highlighting the pitfalls and risks associated with using only binary gender/sex markers, we aim to emphasize the importance of greater precision in defining and operationalizing gender relations and sex-linked biology in the research process. Our goal is to guide future researchers and interpreters of epidemiologic research in adopting best practices for incorporating gender/sex variables in secondary data analysis, thereby enhancing the accuracy, relevance, and applicability of epidemiologic research.

In particular, using data from the National Health and Nutrition Examination Survey (NHANES) – a commonly-used dataset in introductory epidemiology courses to teach basic epidemiologic methods – students will learn to evaluate secondary data to understand what is (and is not) included in variables related to gender/sex. Using NHANES to examine depression in adult respondents, students will dig into the dataset codebooks to understand what dimensions of gender/sex are included in the data – and what dimensions are missing. By engaging with questions of gender/sex multidimensionality alongside a specific research question, students will learn how to evaluate the strengths and limitations of a given secondary dataset in answering questions about the role of gender/sex in shaping the distribution of a given health outcome. Students will learn to identify limitations of using the current proxy variables for gender/sex in secondary databases and consider how imprecise measures can lead to incorrect inferences by

exploring how including other social variables in analysis (here, occupation) with gendered norms may lead to different conclusions.

Learning Objectives for Students:

Through engagement with this teaching example, students will be able to:

- Evaluate the strengths and limitations of a given dataset in pursuing questions related to, or including, gender/sex.
- Describe best practices for using gender/sex variables in common epidemiological methods, thinking through whether gender/sex are causally important in research, including in stratified analyses.
- Understand the limitations of using proxy measures in secondary data sources related to the multidimensions of gender/sex, and its possible consequences for generating accurate effect estimates.

Teaching Methods:

- Lecture Component:

The introductory lecture will briefly cover the key concepts of gender/sex and explore its multidimensional nature, based on an article from Bauer (2023). The lecture will then introduce a Directed Acyclic Graph (DAG) to explore the causal relations of gender/sex on a health outcome, drawing on an article from Krieger (2003). By engaging in critical thinking on the implications of the definition of gender/sex in epidemiology research, students will gain an understanding of the importance of precision when conducting research involving gender/sex.

- Interactive Lab Session:

In the lab, instructors will lead students through data analysis exercises using the NHANES dataset, focusing on depression. Students will examine the relationship between gender/sex, depression, and occupation. They will revisit the multidimensionality of gender/sex (e.g., as described in Bauer, 2023) and be tasked with identifying the most relevant dimensions for their analysis. This includes creating a conceptual DAG to illustrate how gender relations and sex-linked biology influence the relationship between gender/sex and depression, drawing on prior literature for inspiration (e.g., Krieger, 2003). Through this exercise, students will gain a deeper understanding of the critical role gender/sex plays in epidemiologic research.

References

Bauer, G. R. (2023). Sex and Gender Multidimensionality in Epidemiologic Research. *American Journal of Epidemiology*, *192*(1), 122–132. https://doi.org/10.1093/aje/kwac173

Krieger, N. (2003). Genders, sexes, and health: What are the connections--and why does it matter? *International Journal of Epidemiology*, *32*(4), 652–657. https://doi.org/10.1093/ije/dyg156