Curtiss Engstrom is a Ph.D. candidate at the University of Michigan – Ann Arbor. He received his B.S. in Sociology with a Concentration in Analysis and Research and in Political Science with Honors from the University of Wisconsin–Madison. He received his M.S. in Survey and Data Science from the University of Michigan – Ann Arbor. His research focuses on questionnaire design, survey methodology, measurement of sexual and gender minorities, and data quality. His current work examines the data quality and measurement of sexual orientation in surveys and how improving measurement and data quality of sexual orientation research affects estimates of cigarette use and cigarette use-related health outcomes.

Measuring Sexual Orientation in Population-Based Survey Research: Harmonization, Estimation, and Dimensionality

Individuals identifying as a sexual minority are one of the fastest-growing demographic groups in the United States. However, research involving sexual minorities at a national level is limited due to sample size and measurement restrictions. Thus, we need to leverage available statistical methods and adapt them to the measurement of sexual orientation. Understanding which methods work best for identifying sexual minorities is essential for generating accurate, high-quality survey estimates. This is especially important for understanding cigarette-smoking-related outcomes among sexual minorities, as they are at higher risk of cigarette use and, consequently, long-term cigarette-smoking-related health outcomes, including lung cancer. This dissertation evaluates new methods for (1) harmonizing measurements of sexual identity across multiple surveys, (2) merging estimates from multiple surveys to improve the accuracy and precision of health-related outcomes among sexual minorities, and (3) examining the effects that multiple conceptualizations of sexual orientation have on estimates related to cigarette smoking.

The first study uses the 2018 National Health Interview Survey (NHIS) and the 2018 National Survey on Drug Use and Health (NSDUH) to compare two methods of harmonizing different measures of sexual identity. We compare collapsing sexual identity into only categories shared between the NHIS and the NSDUH against using a machine learning approach to bridge an expanded four-category sexual identity measure from the

NHIS onto the NSDUH, which used only a three-category measure, when computing cigarette smoking outcome estimates, including lifetime cigarette use and eligibility for lung cancer screening. Results show the modest impact bridging has on cigarette-smoking-related estimates among subgroups defined by sexual identity and discuss the implications of bridging compared to collapsing for measurement harmonization.

The second study examines two methods for combining cigarette smoking outcome estimates by sexual identity using the harmonized 2018 NHIS and 2018 NSDUH data. We compared the weighted finite-population Bayesian bootstrap (WFPBB) and multilevel regression and poststratification (MRP) and generated estimates by sexual identity in each dataset. We then combined the estimates and compared their accuracy and precision against an external benchmark, the Behavioral Risk Factor Surveillance System, to examine relative bias. We also conducted a simulation study combining NHIS and NSDUH estimates using both methods to examine performance under controlled conditions. Overall, MRP produced more accurate and precise combined estimates than the WFPBB in the empirical analyses.

The third study evaluates the conceptualization of sexual orientation in surveys. Using the 2012 National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) and two collection periods (2017-2019 and 2022-2023) of the National Survey of Family Growth (NSFG), we compared sexual orientation based on sexual identity alone against a composite measure of sexual identity, sexual attraction, and past-year sexual behavior. The composite measure was primarily driven by sexual identity, except for a small group of heterosexual-identified individuals whose sexual attraction and/or behavior does not align with their sexual identity. These results held when stratified by binary sex, adolescent/young adult status, and their interaction. Estimates of disparities related to cigarette smoking outcomes did not differ when using sexual identity alone versus the composite measure, suggesting sexual identity alone is sufficient for these analyses. We discuss the implications for questionnaire design and argue that while sexual identity alone may be sufficient for some analyses, continued collection of all three dimensions of sexual orientation remains important for capturing variation in sexual orientation that sexual identity alone cannot detect.