Application of Behavioral Risk Factor Surveillance System Sampling Weights to Transgender Health Measurement

Ethan C. Cicero, PhD, RN [Postdoctoral Scholar Fellow],
Department of Community Health Systems, University of California, San Francisco, School of Nursing, San Francisco, California, United States

Sari L. Reisner, ScD [Assistant Professor] [Associate Scientific Researcher] [Affiliated Research Scientist and Director],
Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, United States

Division of General Pediatrics, Boston Children's Hospital, Boston, Massachusetts, United States

Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States

Transgender Health Research Team, The Fenway Institute, Fenway Health, Boston, Massachusetts, United States

Elizabeth I. Merwin, PhD, RN, FAAN [Dean],
University of Texas at Arlington, College of Nursing and Health Innovation, Arlington, Texas, United States

Janice C. Humphreys, PhD, RN, FAAN [Professor],
Duke University School of Nursing, Durham, North Carolina, United States.

Susan G. Silva, PhD [Associate Professor]
Duke University, Schools of Nursing and Medicine, Durham, North Carolina, United States

Abstract

Corresponding author: Ethan C. Cicero, PhD, RN, University of California, San Francisco, School of Nursing, Department of Community Health Systems, 2 Koret Way, San Francisco, CA 94143. ethan.cicero@ucsf.edu.
Ethan C. Cicero, PhD, RN, is a Postdoctoral Scholar Fellow, Department of Community Health Systems, University of California, San Francisco, School of Nursing, San Francisco. At the time this research was completed, they were a PhD Candidate, Duke University School of Nursing, Durham, North Carolina.
Sari L. Reisner, ScD, Assistant Professor, Department of Pediatrics, Harvard Medical School Boston, Massachusetts; Associate Scientific Researcher, Division of General Pediatrics, Boston Children's Hospital, Boston, Massachusetts; Assistant Professor, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; Affiliated Research Scientist and Director, Transgender Health Research Team, The Fenway Institute, Fenway Health, Boston, Massachusetts.
Elizabeth I. Merwin, PhD, RN, FAAN, Dean, University of Texas at Arlington, College of Nursing and Health Innovation, Arlington, Texas. At the time this research was completed, she was Vice Dean, Duke University School of Nursing, Durham, North Carolina.
Janice C. Humphreys, PhD, RN, FAAN, is Professor and **Susan G. Silva, PhD**, is Associate Professor, Duke University School of Nursing, Durham, North Carolina. **Susan G. Silva, PhD**, is also Associate Professor, Duke University, School of Medicine, Durham, North Carolina.

Background—Obtaining representative data from the transgender population is fundamental to improving their health and well-being and advancing transgender health research. The addition of the Behavioral Risk Factor Surveillance System (BRFSS) gender identity measure is a promising step towards better understanding transgender health. However, methodological concerns have emerged regarding the validity of data collected from transgender participants and its effect on the accuracy of population parameters derived from those data.

Objectives—To provide rationale substantiating concerns with the formulation and application of the 2015 BRFFS sampling weights and address the methodological challenges that arise when using this surveillance data to study transgender population health.

Methods—We examined the 2015 BRFSS methodology and used the BRFSS data to present a comparison of poor health status using two methodological approaches (a matched-subject design and the full BRFSS sample with sampling weights applied) to compare their effects on parameter estimates.

Results—Measurement error engendered by BRFSS data collection procedures introduced sex/gender identity discordance and contributed to problematic sampling weights. The sex-specific "raking" algorithm used by BRFSS to calculate the sampling weights was contingent on the classification accuracy of transgender by participants. Due to the sex/gender identity discordance of 74% of the transgender women and 66% of transgender men, sampling weights may not be able to adequately remove bias. The application of sampling weights has the potential to result in inaccurate parameter estimates when evaluating factors that may influence transgender health.

Discussion—Generalizations made from the weighted analysis may obscure the need for healthcare policy and clinical interventions aimed to promote health and prevent illness for transgender adults. Methods of public health surveillance and population surveys should be reviewed to help reduce systematic bias and increase the validity of data collected from transgender people.

Keywords

Transgender individuals have an incongruence between the sex assigned to them at birth and their current gender identities; conversely, those with an alignment between their assigned sex and their gender identities are classified as cisgender individuals (Institute of Medicine [IOM], 2011). Members of the transgender population who self-identify as gender nonbinary (GNB) have a gender identity that is as at odds with cultural and social norms (e.g., gender nonconforming, genderqueer, gender expansive, or having a nonbinary gender outside the traditional female–male binary) (James et al., 2016). Transgender women (TW), previously referred to as "male-to-female," are women who were assigned male at birth and currently identity as female/women; transgender men (TM), previously referred to as "female-to-male," are men who were assigned female at birth and currently identify as male/men (IOM, 2011). In this article, we use a person-centered and an affirming approach in describing transgender persons (e.g., TW, TM); however, not all researchers, federal, national, and state agencies have adopted this perspective. The U.S. transgender population comprises 1.4

million TW, TM, and GNB adults; these subgroups are diverse in gender identities, expressions, and roles (Meyer et al., 2017).

Beginning in 2014, the Behavioral Risk Factor Surveillance System (BRFSS) offered all participating states and U.S. territories an optional survey module capturing data on sexual orientation and gender identity (Centers for Disease Control and Prevention [CDC], 2015a). BRFSS is an annual cross-sectional telephone survey conducted by the U.S. CDC and implemented in all states and participating U.S. territories to collect data on health and health behaviors of noninstitutionalized adults ages 18 years or older (CDC, 2015b). Nineteen states and Guam implemented this module in 2014, and in 2015, 22 states included the optional module. With the inclusion of this optional module, BRFSS data represents a first look at the health status of the U.S. transgender population, albeit with limitations. The CDC has suggested that BRFSS data be analyzed with the sampling weights provided to account for the complex design of survey data and increase generalizability of sample results (CDC, 2012). While this is a common approach with probability-based samples, the calculation of BRFSS sample weights relied heavily on sex-based data from the U.S. Census Bureau (CDC, 2015c). Surveys and censuses administered by the Census Bureau (e.g., the American Community Survey [ACS] and the Decennial Census) do not include gender identity measures used to identify transgender people living in the U.S.

Obtaining representative data from the transgender population is fundamental to improving their health and well-being and advancing transgender health research. The addition of the BRFSS gender identity measure is a promising step towards better understanding transgender health. Nonetheless, methodological concerns have emerged regarding the validity of data collected from transgender participants and its effect on the accuracy of population parameters derived from those data. Riley et al. (2017) described how measurement error engendered by BRFSS data collection procedures produced a discordance between surveyor-assigned sex and self-reported gender identity. This discordance introduced misclassification bias which then influenced the administration of sex-specific survey items provided to the 2014 BRFSS transgender participants (Riley et al., 2017). As a result, 30% of the transgender participants were asked inappropriate sex-specific and anatomically based survey items (e.g., TM being asked about prostate cancer screening) (Riley et al., 2017). Their investigation revealed potential error from the sex/gender identity discordance—namely implausible responses to sexual and reproductive health questions.

Objectives

In this report, we build upon the work of Riley et al., (2017) and focus on the methodological issues pertaining to the discordance—particularly how it contributed towards problematic BRFSS sampling weights. In subsequent sections, we describe the 2015 BRFSS methodology—including the underpinnings for the BRFSS sampling weights—and provide rationale substantiating concerns with the formulation and application of the 2015 BRFSS sampling weights; publicly available data that are being used in current and new transgender health research (Blosnich et al., 2017; Crissman et al., 2019; Downing & Przedworski, 2018; Lagos, 2018; Seelman et al., 2018; Streed et al., 2017). By using data from 22 states that adopted the gender identity measure in 2015, we present a comparison of

poor health status using two methodological approaches (a matched-subject design without BRFSS sampling weights, and the complete BRFSS sample with sampling weights applied) to compare the effect of the two approaches on parameter estimates. Finally, we discuss methodological and conceptual factors for consideration by researchers and federal, national, and state agencies when collecting population-based data for transgender people.

Measurement Error and Misclassification Bias

Following BRFSS sample design and data collection standards, states obtained a probability sample of all households with landline and cellular telephones (CDC, 2016). Although it is beyond the scope of this article to describe BRFSS design weights and sampling procedures, BRFSS technical documentation are publicly available online (CDC, 2015). BRFSS interviews were conducted at the state level by health personnel or contractors, and the BRFSS questionnaire was administered to participants using both landlines and cellular telephones (CDC, 2016).

The BRFSS interview began with a core set of standardized questions and was followed by optional modules and then any state-added questions. For the module used to identify transgender participants, participants were asked, "Do you consider yourself to be transgender?", and if affirmed they were asked, "Do you consider yourself to be male-to-female, female-to-male, or gender nonconforming?" (CDC, 2014, p. 69). The gender identity module identified 369 TW, 239 TM, and 156 GNB adults from 22 states.

BRFSS participants were not asked their sex, instead surveyors determined it by the participant's vocal timbre (CDC, 2014). By comparing the gender identity (TW, TM, and GNB) of participants who self-identified as transgender and the sex assigned to them by surveyors, we discovered measurement error that produced misclassification bias. Subsequently, 74% of TW and 66% of TM had a discordance between surveyor-assigned sex and self-reported gender identity (Table 1). It is not possible to determine the classification accuracy for the GNB participants because, per the BRFSS definition, these individuals do not identity as a man or woman. This measurement error and misclassification bias that ensued influenced the sex-specific algorithm used to create BRFSS sampling weights (CDC, 2016).

Sampling Weight Method

In 2011, BRFSS began using a weighting method known as *iterative proportional fitting*, or *raking*, that allowed for the incorporation of additional demographic variables into the statistical weighting process (CDC, 2012). Through a series of data processing-intensive iterations, raking adjusted each participant's sampling weight based on ranking control variables, or margins, until weighted frequencies and population frequencies for each margin converged (CDC, 2012). BRFSS data were raked by categories of age by sex, race and ethnicity, education, marital status, regions or counties within states, sex by race and ethnicity, telephone source, renter or owner status, and age groups by race and ethnicity (CDC, 2016).

Raking required the use of auxiliary population information. BRFSS raked the sample data using county-level intercensal population estimates from the Nielsen Company that were derived from the Decennial Census, and sex-specific data on education, marital status, and tenure from the 2009–2013 ACS (CDC, 2015c). The Census Bureau does not include measures used to identify transgender participants. As a result, data from the Census Bureau represent gender binary, only accounting for the lives of cisgender men and cisgender women.

The CDC has suggested that BRFSS data be analyzed with the sampling weights provided to increase the external validity and generalizability of the sample results (CDC, 2012). However, the sex-specific raking algorithm used was contingent on the classification accuracy of transgender BRFSS participants. Data from the 74% of TW who were assigned male by surveyors had their data raked using population-level data for cisgender males, and population-level data for cisgender females were used to rake the 66% of TM who were assigned female by surveyors. Measurement error and the sex/gender identity discordance contributed to inaccurate sampling weights for the transgender participants. For this reason, BRFSS sampling weights may not be able to adequately remove bias in the sample. Resulting parameter point and interval estimates derived from the weighted data may be affected. The sex misclassification bias has the potential to affect the validity and precision of statistical conclusions regarding transgender people.

Addressing Problematic Sampling Weights

Alternative methodological approaches are needed to address the BRFSS sampling weights. Correctly classifying all transgender participants based on their gender identities and applying the BRFSS ranking algorithm to derive new sampling weights is not currently possible, for at least two reasons. First, several of the ranking margins require county-level data. However, in order to protect the confidentiality of BRFSS participants, this substate geographic identifier was suppressed by BRFSS for states that employed disproportionate sampling methods (21 out of the 22 states that used the gender identity module in 2015) (CDC, 2016). Second, GNB participants have a gender identity outside the female-male binary, and arbitrarily imposing a binary gender distinction to GNB participants in order to rake their data would potentially (re)introduce measurement bias, which may affect parameter estimates.

Modifying a case-control study design—an approach used to investigate a dichotomous disease and/or exposure variable—provides one way to account for the drawbacks when using BRFSS sampling weights. In traditional case-control studies, cases are study subjects that were exposed to or have the condition under study, and controls are subjects who lack the condition and are also a representative sample of the cases selected. Controls are matched to cases by extraneous variables known a priori to be correlated with the condition under study (Kupper et al., 1981). When exploring differences between transgender ("case") and cisgender ("control") participants, a matched-subject design can create a more balanced sample that mitigates the effects of confounders, increases the accuracy of parameter estimates, and addresses the systematic bias attributed to BRFSS sampling weights—but it does not correct the inaccurate sampling weights (Kupper et al., 1981).

Properly matching transgender cases with cisgender controls is a key consideration for a matched-subject design; it is also a challenging task. Currently, there are no available data suggesting that the health of a transgender adult is equivalent to that of a cisgender adult with the same sex assignment (e.g., TM: cisgender woman). One option is to match a single transgender case with two cisgender controls: one cisgender man (CM) and one cisgender woman (CW) (Reisner et al., 2014). Taylor (1986, p. 29) recommends not using more than four controls, as the "marginal return rapidly diminishes as the number of controls per case increases." Choosing the optimal variables for the matching process is critical. Conceptually, it is important to consider the study aims in terms of identifying an appropriate matching algorithm.

Methods

Using data from the 22 states that adopted the 2015 BRFSS gender identity module, we present a comparison of poor health status using two methodological approaches to illustrate their differential effects on parameter estimates. The first approach (matched analysis) utilized a matched-subject design without the use of BRFSS sampling weights, while the second approach (weighted analysis) included all transgender and cisgender participants, accounted for BRFSS's complex survey design, and included the sampling weights. This secondary analysis of publicly available data was reviewed and received an exemption by the institutional review board at the Duke University School of Nursing.

The goal of this illustration was to estimate the overall prevalence of poor health status, present the adjusted odds ratios (aOR) with their 95% confidence intervals/limits (CI/L) for each of the five study groups (TW, TM, GNB, CM, CW) using the two approaches, and then compare the pattern of results from these two methods. Poor health status was derived from a single BRFSS item that asked participants to rate their general health as *excellent, very good, good, fair,* or *poor.* From this, a dichotomous poor health status variable was created and coded as 0 (excellent, very good, or good) or 1 (fair or poor).

Analysis Sample

Two analysis samples were utilized. Based on the participant's response to the gender identity measure, transgender participants were classified as TW, TM, or GNB adults. For those participants who indicated they were not transgender, surveyor-assigned sex was used to classify the participant as CM or CW. These identity groups represent five distinct levels of study group. The weighted analysis sample included all transgender and cisgender participants from the 22 states. A matched-subject design was used to obtain a 1:4 matched analysis sample in which each transgender participant was matched with two CM and two CW. Matching variables were state, metropolitan status area, age (partitioned into 5-year increments, 18–24 to 70–74, > 80), and race/ethnicity (White, non-Hispanic; Black, non-Hispanic; Hispanic; and any other race/ethnicity). These variables were selected based on their known influence on health outcomes in stigmatized populations (Cagney et al., 2005; Ferraro et al., 2016; Link & Phelan, 1995; Phelan et al., 2010). For this illustration, we intentionally did not match on other sociodemographic variables because of known social

and economic inequities between cisgender and transgender populations and among transgender subgroups (Cicero et al., 2019; James et al., 2016).

Statistical Plan

A logistic regression model was fit for each analysis sample to determine whether study group predicted poor health status, after covarying for age, race, and socioeconomic variables known to influence health outcomes for stigmatized groups (educational attainment, employment status, and health insurance) (Link & Hatzenbuehler, 2016; Link & Phelan, 1995). Using guidance from BRFSS, one logistic regression model (weighted analysis) was estimated by applying BRFSS sampling weights and accounting for the complex survey design (CDC, 2016). The second logistic regression model (matched analysis) did not include BRFSS sampling weights or account for the complex survey design. Within these models, study group does not have a single reference group, rather a priori pairwise comparisons of the groups were conducted to compare the aORs with their CI/Ls for poor health among the groups. Annual household income was excluded as a covariate due to the missing rate per study group exceeding 12%.

Nondirectional statistical tests were performed using SAS 9.4® with the level of significance set at 0.05 for the overall study group and covariate effects. To further evaluate differences in poor health status among the five study groups, a priori pairwise comparisons of the groups were conducted. If the overall Wald chi-square value for study group from the logistic regression was statistically significant, two-sided multiple comparisons were conducted using a closed test procedure with significance set at 0.05 for each test. If the overall study group result was not significant, a more conservative approach—the Benjamini–Hochberg method—was planned to safeguard against a Type I error (Koch & Gansky, 1996). This method controls the false discovery rate using sequential modified Bonferroni correction for multiple hypothesis testing (Benjamini & Hochberg, 1995).

The overall missing rate for poor health status ranged from 0.0%–0.4% per group. With covariates incorporated in the model, the matched analysis included 98.0% of the matched-subject sample and the weighted analysis utilized 98.6% of the BRFSS sample. The missing rate per group ranged from 1.4% (CW, weighted) to 3.8% (TM, both samples). Since the missing rate was less than 4% per group, we elected to not apply imputation methods.

Results

Descriptive statistics for demographic characteristics, covariates in the regression models, and poor health status unadjusted for covariates are summarized for each analysis in Table 2. The effect of sampling weights on these summary statistics can be observed in this table, particularly among the transgender subgroups in which the sample size per group did not change based on analysis sample. When sampling weights were applied, the mean age decreased by nearly 10 years for each transgender subgroup. Sampling weights had a considerable effect on TM and TW data. With weights applied, TM had over a 40% increase in those who have at least a high school education and a 30% reduction in poor health. For those who reported being uninsured, sampling weights produced nearly a 130% increase for TW and 175% increase for TM as compared to their respective unweighted percentages.

Table 3 provides the logistic regression results for both analysis samples. Study group was a predictor of poor health status, after adjusting for covariates in both analysis samples. The covariate results were similar for both analysis samples—with the exception of health insurance. The likelihood of poor health status was associated with being older, a racial/ethnic minority, not working, and having a high school education or less. Not having health insurance, however, was associated with poor health status in the weighted analysis only.

The overall effect of study group was significant for both analysis samples. Figure 1 presents the results for the study group contrasts for the poor health outcome, and the aORs with their 95% CI/L for each pairwise contrast. Among three pairwise transgender subgroup comparisons, the matched and weighted analysis samples both indicated no difference between TW and TM. Relative to GNB adults, the matched analysis sample revealed that both TW (*aOR* = 0.65, 95% CI [0.42, 1.01]) and TM (aOR = 0.54, 95% CI [0.33, 0.88]) had lower odds of poor health status; whereas, only TM (*aOR* = 0.33, 95% CL [0.12, 0.91]) had lower odds of poor health status in the weighted analysis.

The contrasts between the transgender and cisgender subgroups in the matched analysis sample indicated that GNB adults were more likely to have poor health status than CM (*aOR* = 1.87, 95% CI [1.27, 2.75]) and CW (*aOR* = 2.10, 95% CI [1.43, 3.09]); additionally, TW were more likely to have poor health status relative to CW (*aOR* = 1.36, 95% CI [1.02, 1.81]). The weighted analysis sample revealed only one significant contrast: CM were more likely to have poor health status (a*OR* = 1.06, 95% CL [1.01, 1.12]) than CW.

The width of the CI increased when BRFSS sampling weights were applied—with the exception of the CM/CW comparison. Given that the transgender subgroups consisted of the same transgender BRFSS participants across both analysis samples, the increase in CIs and variation in the aORs were particularly notable for the transgender subgroup comparisons.

Discussion

The BRFSS optional gender identity module provides researchers an unprecedented opportunity to conduct comparative studies among transgender communities and with their cisgender counterparts. Nevertheless, it is important to consider the limitations of BRFSS methods, including those recommended for data analysis, particularly if the goal is to achieve estimates that more accurately represent the population from which the sample was drawn.

Our findings suggest that the use of BRFSS sampling weights can result in biased parameter estimates when evaluating factors that may influence transgender health. As expected, narrower CIs were generated by using a matched-subject design versus the weighted analysis sample. However, the effect of BRFSS sampling weights is particularly evident in the pairwise comparisons of transgender subgroups, as both analysis samples included the same transgender participants. In these contrasts, the application of BRFSS sampling weights contributed towards not only inaccurate poor health status estimates, but also less precision in the parameter intervals. By using a matched-subject design, the CIs were narrower, thus shifting the aORs; causing the aORs to be discrepant across analysis samples.

The CDC implemented changes to the 2016 BRFSS methodology for data collection on sex. BRFSS surveyors will now ask participants, "Are you male or female" (CDC, 2015a). However, the 2014 and 2015 BRFSS data remain publicly available, and these data and the CDC-provided BRFSS sampling weights are being used in current and new research studies (Blosnich et al., 2017; Crissman et al., 2019; Downing & Przedworski, 2018; Lagos, 2018; Seelman et al., 2018; Streed et al., 2017). Given the discordance between surveyor-assigned sex and self-reported gender identity that occurred in the 2014 and 2015 BRFSS, both datasets include problematic BRFSS sampling weights. Therefore, results from analyses using BRFSS sampling weights with either dataset may lead to inaccurate generalizations.

Conventional approaches to analyzing probability-based, complex survey data may affect the validity and precision of statistical conclusions regarding transgender people. A matched-subject design provided a better indication between being transgender and having poor health— above and beyond the covariates that are known to influence health. It provided more accurate parameter point estimates, higher precision in parameter interval estimates, and group differences were more apparent. The approach utilized to create our matched-subject analysis sample will help researchers match cisgender controls to transgender cases in studies using BRFSS data and in future studies that explore individuals who do not identify within binary distinctions of gender.

Limitations

Limitations of our findings include the exclusive focus on poor health, included covariates, and the approach taken to establish a matched-subject analysis sample. However, to our knowledge, the illustrative example discussed here represents the first published pairwise comparisons of poor health status among transgender subpopulations and in relation to their cisgender male and female counterparts. Both selection bias and the limited number of states that used the BRFSS gender identity module also contribute to limitations to the current study.

Research Implications

Transgender health research is hindered by the lack of gender identity measures in health-related research and data collection efforts by the Census Bureau. Without counting the U.S. transgender population, the Census Bureau fails to fulfill their mission: "To serve as the nation's leading provider of quality data about its people and economy" (United States Census Bureau, n.d.). Moreover, their approach has downstream effects in how federal, national, and state agencies conduct health surveillance, particularly in sampling approaches and calculating sampling weights. To accurately calculate sampling weights, transgender research participants should be classified based on their gender identities or gender, and not their sex assigned at birth. This approach—grounded in a gender binary and influenced by the Census Bureau's data collection efforts—will only address the misclassification of transgender women and transgender men. Methods are still needed that account for persons who self-identify as GNB, as their identities fall outside the female-male binary, and arbitrarily imposing a binary gender distinction to GNB participants in order to calculate their sampling weight, would potentially introduce measurement bias, which may affect parameter estimates.

Data collection approaches are needed that routinely include transgender-inclusive measures by all researchers, federal, national, and state agencies. The Gender Identity in U.S. Surveillance (2014) group recommends using a two-step method for identifying transgender people, asking assigned sex and current gender identity. For example: (a) "How do you describe yourself?" (male, female, TW, TM, GNB, a different identity); and (b) "What sex were you assigned at birth, on your original birth certificate?" (male or female). By capturing both current gender identity and sex assignment, the two-step approach can minimize misclassification of transgender people, identify transgender respondents who may not use the term transgender when describing themselves, and be more effective in gathering accurate and useful information to improve the health of transgender people.

The field of transgender health research represents an untapped and important area of focus for the nursing discipline. The current body of health-related literature does not reflect the nursing's holistic perspective on health and healthcare; the "health or wholeness of human beings as they interact with their environment" with both wellness and illness comprising health (Donaldson & Crowley, 1978, p. 113). As nurse scientists and researchers study various aspects of transgender health, including the adoption of transgender-inclusive measures, they may benefit from collaborating with researchers who specialize in methods for transgender and gender health research. Through such collaborations, nurse scientists and researchers can contribute high-quality research grounded by the nursing metaparadigm; research that has the potential to extend and broaden the current knowledge of transgender health. The knowledge gained by nursing research can identify and prioritize health-related needs and provide the foundation to develop clinical interventions aimed at promoting health and preventing illness for transgender and gender nonbinary people.

Conclusion

When exploring transgender health, the use of BRFSS sampling weights may inadvertently portray a biased depiction of the health of transgender adults in the U.S. Generalizations made from the weighted analysis may obscure the need for healthcare policy and clinical interventions aimed to promote health and prevent illness for transgender adults—particularly TW and TM. Additional implications may surface in analyses using BRFSS data when exploring other health outcomes and associations with socioeconomic characteristics. Methods of current and future public health surveillance and population surveys should be reviewed to help reduce systematic bias and to increase the validity of data collected from diverse transgender people.

Acknowledgments

References

Benjamini Y, & Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57, 289–300. 10.1111/j.2517-6161.1995.tb02031.x

Blosnich JR, Lehavot K, Glass JE, & Williams EC (2017). Differences in alcohol use and alcohol-related health care among transgender and nontransgender adults: Findings from the 2014 Behavioral Risk Factor Surveillance System. Journal of Studies on Alcohol and Drugs, 78, 861–866. 10.15288/jsad.2017.78.861 [PubMed: 29087820]

Cagney KA, Browning CR, & Wen M (2005). Racial disparities in self-rated health at older ages: What difference does the neighborhood make? Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 60, S181–S190. 10.1093/geronb/60.4.s181

Centers for Disease Control and Prevention. (2012). Methodologic changes in the Behavioral Risk Factor Surveillance System in 2011 and potential effects on prevalence estimates. MMWR: Morbidity and Mortality Weekly Report, 61, 410. [PubMed: 22672976]

Centers for Disease Control and Prevention. (2014, 12 29). 2015 Behavioral Risk Factor Surveillance System questionnaire. Retrieved from https://www.cdc.gov/brfss/questionnaires/pdf-ques/2015-brfss-questionnaire-12-29-14.pdf July 1, 2018

Center for Disease Control and Prevention. (2015b). 2015 BRFSS data and documentation: 2015 data information. Retrived from https://www.cdc.gov/brfss/annual_data/annual_2015.html

Centers for Disease Control and Prevention. (2015a, 14 10). 2016 Behavioral Risk Factor Surveillance System questionnaire. Retrieved from https://www.cdc.gov/brfss/questionnaires/pdf-ques/2016_BRFSS_Questionnaire_FINAL.pdf 71, 2018

Centers for Disease Control and Prevention. (2015b, 9). Behavioral Risk Factor Surveillance System: Overview: BRFSS 2014. Retrieved from https://www.cdc.gov/brfss/annual_data/2014/pdf/overview_2014.pdf

Centers for Disease Control and Prevention. (2015c). Behavioral Risk Factor Surveillance System weighting BRFSS data 2015. Retrieved from https://www.cdc.gov/brfss/annual_data/2015/pdf/weighting_the-data_webpage_content.pdf July 1, 2018

Centers for Disease Control and Prevention. (2016, 7). Behavioral Risk Factor Surveillance System: Overview: BRFSS 2015. Retrieved from https://www.cdc.gov/brfss/annual_data/2015/pdf/overview_2015.pdf

Cicero EC, Reisner SL, Silva SG, Merwin EI, & Humphreys JC (2019). Health care experiences of transgender adults: An integrated mixed research literature review. Advances in Nursing Science, 42, 123–138. 10.1097/ans.0000000000000256 [PubMed: 30839332]

Crissman HP, Stroumsa D, Kobernik EK, & Berger MB (2019). Gender and frequent mental distress: Comparing transgender and non-transgender individuals' self-rated mental health. Journal of Women's Health, 28, 143–151. 10.1089/jwh.2018.7411

Donaldson SK, & Crowley DM (1978). The discipline of nursing. Nursing Outlook, 26, 113–120. [PubMed: 245616]

Downing JM, & Przedworski JM (2018). Health of transgender adults in the U.S., 2014–2016. American Journal of Preventive Medicine, 55, 336–344. 10.1016/j.amepre.2018.04.045 [PubMed: 30031640]

Ferraro KF, Schafer MH, & Wilkinson LR (2016). Childhood disadvantage and health problems in middle and later life: Early imprints on physical health? American Sociological Review, 81, 107–133. 10.1177/0003122415619617 [PubMed: 27445413]

Gender Identity in U.S. Surveillance Group. (2014, 9). Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys. The Williams Institute Retrieved from https://williamsinstitute.law.ucla.edu/wp-content/uploads/geniuss-report-sep-2014.pdf July1 2018

Institute of Medicine. (2011). The health of lesbian, gay, bisexual, and transgender people: Building a foundation for better understanding. The National Academies Press 10.17226/13128

James SE, Herman JL, Rankin S, Keisling M, Mottet L, & Ana M (2016). The report of the 2015 U.S. transgender survey. National Center for Transgender Equality https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf

Koch GG, & Gansky SA (1996). Statistical considerations for multiplicity in confirmatory protocols. Drug Information Journal, 30, 523–534. 10.1177/009286159603000228

Kupper LL, Karon JM, Kleinbaum DG, Morgenstern H, & Lewis DK (1981). Matching in epidemiologic studies: Validity and efficiency considerations. Biometrics, 37, 271–291. 10.2307/2530417 [PubMed: 7272415]

Lagos D (2018). Looking at population health beyond "male" and "female": Implications of transgender identity and gender nonconformity for population health. Demography, 55, 2097–2117. 10.1007/s13524-018-0714-3 [PubMed: 30255426]

Link B, & Hatzenbuehler ML (2016). Stigma as an unrecognized determinant of population health: Research and policy implications. Journal of Health Politics, Policy and Law, 41, 653–673. 10.1215/03616878-3620869

Link BG, & Phelan J (1995). Social conditions as fundamental causes of disease. Journal of Health and Social Behavior, 80–94. 10.2307/2626958 [PubMed: 7560851]

Meyer IH, Brown TNT, Herman JL, Reisner SL, & Bockting WO (2017). Demographic characteristics and health status of transgender adults in select US regions: Behavioral Risk Factor Surveillance System, 2014. American Journal of Public Health, 107, 582–589. 10.2105/AJPH.2016.303648 [PubMed: 28207334]

Phelan JC, Link BG, & Tehranifar P (2010). Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. Journal of Health and Social Behavior, 51, S28–S40. 10.1177/0022146510383498 [PubMed: 20943581]

Reisner SL, White JM, Bradford JB, & Mimiaga MJ (2014). Transgender health disparities: Comparing full cohort and nested matched-pair study designs in a community health center. LGBT Health, 1, 177–184. 10.1089/lgbt.2014.0009 [PubMed: 25379511]

Riley NC, Blosnich JR, Bear TM, & Reisner SL (2017). Vocal timbre and the classification of respondent sex in US phone-based surveys. American Journal of Public Health, 107, 1290–1294. 10.2105/ajph.2017.303834 [PubMed: 28640682]

Seelman KL, Miller JF, Fawcett ZER, & Cline L (2018). Do transgender men have equal access to health care and engagement in preventive health behaviors compared to cisgender adults? Social Work in Health Care, 57, 502–525. 10.1080/00981389.2018.1462292 [PubMed: 29708468]

Streed CG Jr., McCarthy EP, & Haas JS (2017). Association between gender minority status and self-reported physical and mental health in the United States. JAMA Internal Medicine, 177, 1210–1212. 10.1001/jamainternmed.2017.1460 [PubMed: 28558100]

Taylor JMG (1986). Choosing the number of controls in a matched case-control study, some sample-size, power and efficiency considerations. Statistics in Medicine, 5, 29–36. 10.1002/sim.4780050106 [PubMed: 3961313]

United States Census Bureau. (n.d.) What we do: Our mission. Retrieved July 18, 2018, from https://www.census.gov/about/what.html#par_textimage
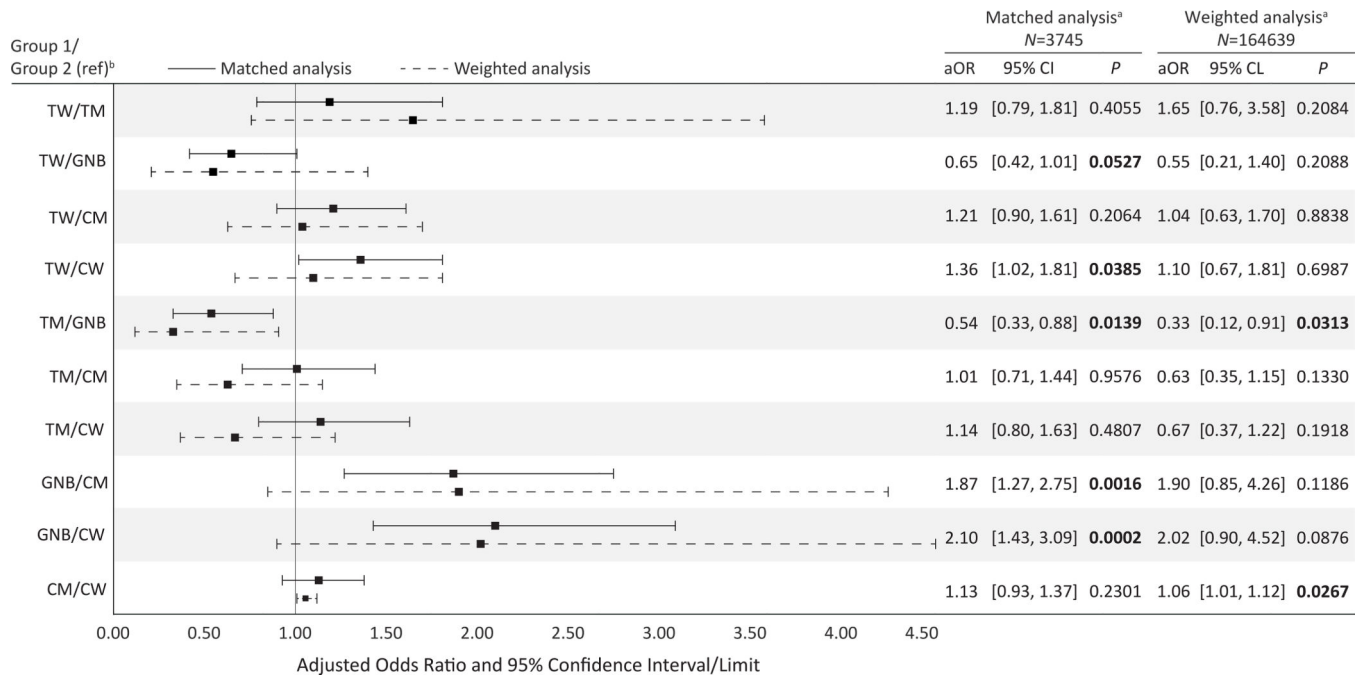
Group 1/ Group 2 (ref)[b]	Matched analysis[a] N=3745			Weighted analysis[a] N=164639		
	aOR	95% CI	P	aOR	95% CL	P
TW/TM	1.19	[0.79, 1.81]	0.4055	1.65	[0.76, 3.58]	0.2084
TW/GNB	0.65	[0.42, 1.01]	**0.0527**	0.55	[0.21, 1.40]	0.2088
TW/CM	1.21	[0.90, 1.61]	0.2064	1.04	[0.63, 1.70]	0.8838
TW/CW	1.36	[1.02, 1.81]	**0.0385**	1.10	[0.67, 1.81]	0.6987
TM/GNB	0.54	[0.33, 0.88]	**0.0139**	0.33	[0.12, 0.91]	**0.0313**
TM/CM	1.01	[0.71, 1.44]	0.9576	0.63	[0.35, 1.15]	0.1330
TM/CW	1.14	[0.80, 1.63]	0.4807	0.67	[0.37, 1.22]	0.1918
GNB/CM	1.87	[1.27, 2.75]	**0.0016**	1.90	[0.85, 4.26]	0.1186
GNB/CW	2.10	[1.43, 3.09]	**0.0002**	2.02	[0.90, 4.52]	0.0876
CM/CW	1.13	[0.93, 1.37]	0.2301	1.06	[1.01, 1.12]	**0.0267**

Legend: —— Matched analysis   - - - - Weighted analysis

X-axis: Adjusted Odds Ratio and 95% Confidence Interval/Limit (0.00, 0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00, 4.50)

Figure 1.

Multivariable Logistic Regression Models: Study Group Pairwise Comparisons of Poor Health Status.

Note. Unweighted analysis: *p*-value from the logistic regression. Weighted analysis: *p*-value from the logistic regression that accounted for complex survey design and sampling weights. Bold indicates statistical significance.

[a] Both analyses included all and only the transgender participants from the 22 states that included the gender identity module.

[b] Group 1/Group 2: pairwise study group comparisons listed on the left side of the figure. Group 2 is the reference group. *aOR* = adjusted odds ratio; CM = cisgender men; CW = cisgender women; CI/L = confidence interval/limit; GNB = gender nonbinary adults; ref = reference; TM = transgender men; TW = transgender women.

Table 1

2015 BRFSS Sex Classification by Vocal Timbre

BRFSS surveyor assignment	TW (*n*=369)		TM (*n*=239)		GNB (*n*=156)	
Male	272	(73.7%)[a]	82	(34.3%)	79	(50.6%)
Female	97	(26.3%)	157	(65.7%)[b]	77	(49.4%)

Note. n (%). Data represents the transgender-identified 2015 BRFSS participants.

[a]Unadjusted number (*n*) and percent (%) of transgender women who were classified as male.

[b]Unadjusted number (*n*) and percent (%) of transgender men who were classified as female. GNB = gender nonbinary adults; TM = transgender men;TW = transgender women.

Table 2

2015 BRFSS Demographics Characteristics and Poor Health Status for the Matched And Weighted Analysis Samples

	Matched analysis (N=3820)[a]					Weighted analysis (N=167059)[b]				
	TW (n=369)	TM (n=239)	GNB (n=156)	CM (n=1528)	CW (n=1528)	TW (n=369)	TM (n=239)	GNB (n=156)	CM (n=70123)	CW (n=96172)
Age	369	239	156	1528	1528	369	239	156	70123	96172
Mean (SEM)	53.89 (0.89)	54.55 (1.15)	52.77 (1.53)	53.88 (0.45)	53.92 (0.45)	46.03 (1.56)	43.26 (1.73)	43.20 (3.10)	47.22 (0.12)	49.03 (0.12)
Race/ethnicity	369	239	156	1528	1528	369	239	156	70123	96172
Racial/ethnic minority[c]	106 (28.73)	73 (30.54)	52 (33.33)	461 (30.17)	462 (30.24)	106 (40.08)	73 (50.56)	52 (48.10)	14,604 (31.60)	20,431 (31.71)
White, non-Hispanic	263 (71.27)	166 (69.46)	104 (66.67)	1067 (69.83)	1066 (69.76)	263 (59.92)	166 (49.44)	104 (51.89)	55519 (68.40)	75741 (68.29)
Educational attainment	368	238	156	1523	1521	368	238	156	69912	95917
High school/GED graduate or less	186 (50.54)	123 (51.68)	74 (47.44)	601 (39.46)	557 (36.62)	186 (59.54)	123 (74.70)	74 (45.72)	23909 (43.52)	32848 (40.65)
Post-secondary school	182 (49.46)	115 (48.32)	82 (52.56)	922 (60.54)	964 (63.38)	182 (40.46)	115 (25.30)	82 (54.28)	46003 (56.48)	63069 (59.35)
Employment status	366	234	154	1514	1516	366	234	154	69692	95601
Not working[d]	205 (56.01)	129 (55.13)	89 (57.79)	681 (44.98)	812 (53.56)	205 (51.37)	129 (46.51)	89 (52.33)	29869 (34.91)	52765 (49.35)
Working	161 (43.99)	105 (44.87)	65 (42.21)	833 (55.02)	704 (46.44)	161 (48.63)	105 (53.49)	65 (47.67)	39823 (65.09)	42836 (50.65)
Annual household income	318	203	135	1326	1259	318	203	135	61581	79185
<$20K	82 (25.79)	43 (21.18)	34 (25.19)	219 (16.52)	245 (19.46)	82 (29.53)	43 (15.82)	34 (22.40)	7727 (14.17)	14336 (19.38)
$20K to <$50K	114 (35.85)	91 (44.83)	50 (37.04)	452 (34.09)	448 (35.58)	114 (39.03)	91 (56.77)	50 (30.88)	19554 (32.54)	27704 (34.11)
$50K	122 (38.36)	69 (33.99)	51 (37.78)	655 (49.40)	566 (44.96)	122 (31.45)	69 (27.41)	51 (46.70)	34300 (53.29)	37145 (46.51)
Health insurance status	363	237	154	1520	1525	363	237	154	69825	95901
Uninsured	28 (7.71)	32 (13.50)	19 (12.34)	150 (9.87)	129 (8.46)	28 (17.59)	32 (37.10)	19 (14.14)	5388 (12.48)	5350 (9.75)
Insured	335 (92.29)	205 (86.50)	135 (87.66)	1370 (90.13)	1396 (91.54)	335 (82.41)	205 (62.90)	135 (85.86)	64437 (87.52)	90551 (90.25)
Poor health status	369	238	156	1525	1525	369	238	156	69918	95888
Poor health	86 (23.31)	50 (21.01)	50 (32.05)	280 (18.36)	269 (17.64)	86 (20.40)	50 (14.66)	50 (30.01)	11927 (16.08)	17431 (17.42)
Good health	283 (76.69)	188 (78.99)	106 (67.95)	1245 (81.64)	1256 (82.36)	283 (79.60)	188 (85.34)	106 (69.99)	57991 (83.92)	78457 (82.58)

Note. Number of participants in study group with data available (*N*) reported for each variable. Mean and standard error of the mean for age presented. For categorical variables, the number (*n*) and percent (%) of *n* are provided; the actual *n* and the weighted percent are reported for the weighted analysis sample. BRFSS= Behavioral Risk Factor Surveillance System; CM = cisgender men; CW = cisgender women; GNB = gender nonbinary adults; SEM = standard error of the mean; TM = transgender men; TW = transgender women.

[a] Each transgender participant was matched with 2 CM and 2 CW by state, metropolitan status area, age, and race/ethnicity.

[b] All participants from the 22 states that included the gender identity module were included in the weighted analysis sample.

[c] Racial/ethnic minority = participants who did not report being white, non-Hispanic.

[d] Not working = homemaker, retired, student, out of work, or unable to work.

Table 3

Logistic Regression Results for Poor Health Outcome

Analytic model	Matched analysis (N=3745)				Weighted analysis (N=164639)			
	χ^2	p	aOR	95% CI	F	p	aOR	95% CL
Study group[a]	16.33	0.0026			2.41	0.0471		
Age	17.35	<0.0001	1.01	[1.01, 1.02]	488.5	<0.0001	1.02	[1.02, 1.02]
Race/ethnicity								
Racial/ethnic minorities	12.98	0.0003	1.41	[1.17, 1.71]	266.8	<0.0001	1.66	[1.57, 1.77]
White, non-Hispanic (ref)								
Educational attainment								
High school/GED graduate or less	35.02	<0.0001	1.69	[1.42, 2.01]	673.8	<0.0001	2.00	[1.90, 2.11]
Post-secondary school (ref)								
Employment status								
Not working[b]	123.05	<0.0001	3.07	[2.52, 3.74]	955.0	<0.0001	2.58	[2.43, 2.74]
Working (ref)								
Health insurance status								
Uninsured	1.58	0.2086	1.21	[0.90, 1.62]	44.8	<0.0001	1.42	[1.28, 1.57]
Insured (ref)								

Note. Matched analysis: χ^2 = Wald chi-square and *P*-value from the logistic regression. Weighted analysis: *F*- and *p*-values from the logistic regression that accounted for complex survey design and BRFSS sampling weights.

[a] Study group includes five categories: transgender women, transgender men, gender nonbinary adults, cisgender men, and cisgender women. Study group does not have a reference group, rather *a priori* pairwise comparisons of the groups are provided in Figure 1.

[b] Not working=homemaker, retired, student, out of work, or unable to work. aOR = adjusted odds ratio; CI/L = confidence interval/limit.