Erasure and Health Equity Implications of Using Binary Male/Female Categories in Sexual Health Research and Human Immunodeficiency Virus/Sexually Transmitted Infection Surveillance: Recommendations for Transgender-Inclusive Data Collection and Reporting

Diana M. Tordoff, MPH, * Brian Minalga, MSW, † Bennie Beck Gross, BA, ‡
Aleks Martin, MSW, SUDP, § Billy Caracciolo, BA, ‡ Lindley A. Barbee, MD, MPH, § ¶
Jennifer E. Balkus, PhD, MPH, * and Christine M. Khosropour, PhD, MPH*

In this commentary, we discuss the health equity implications of collecting and reporting data stratified by binary male/female categories in sexual health and sexually transmitted infection (STI)/human immunodeficiency virus (HIV) research and surveillance without meaningful inclusion of transgender, nonbinary, and gender diverse identities. Numerous institutional bodies have recommended collecting and reporting data disaggregated by trans-inclusive categories, including the National Institutes of Health and the National Academy of Sciences.1,2 However, a review of articles published by Sexually Transmitted Diseases in 2020, for example, yielded 121 articles (of 166 that included human participants) that only reported data by male/female categories and conflated sex assigned at birth with gender (eg, by referring to people assigned female at birth as women, and people assigned male at birth as men). In addition, core surveillance systems and large national data sources, such as the Centers for Disease Control and Prevention 2018 STD Surveillance Report and Health People 2030, also continue to report data in this way.

Trans adults comprise 0.6% of the population, or at least 1.4 million adults living in the United States, and a recent study estimated that approximately 1.2 million adults in the US identify as nonbinary.3-5 In addition, an estimated 1.3% to 2.7% of adolescents in the United States identify as trans or nonbinary,6-8 Trans and nonbinary people are disproportionately burdened by poor health outcomes, including STIs and HIV, because of sociostructural factors rooted in stigma and discrimination.9-11 In addition, trans and nonbinary people report significant barriers to health care, including harassment, refusal of care by health care providers, and difficulty obtaining specific and trans-affirming health information.12,13 Notably, there are significant disparities within trans populations; for example, Black and Indigenous trans women are disproportionately living with HIV and experience syndemic conditions (such as unstable housing and reliance on sex work because of economic disenfranchisement) because of the intersectional impact of structural racism, colonialism, and misogyny, as well as transphobia.14,15 Lastly, although this commentary focuses on trans-inclusivity, it is essential to acknowledge the harms that are inflicted upon intersex individuals (who may or may not identify as transgender) through the use of the same prescriptive male/female sex categories in medicine and research, which are a product of Western colonial thought.16 Because of sociocultural and legal mandates in the United States that categorize individuals as either male or female, intersex infants and children are often unnecessarily and involuntarily subjected to invasive surgeries, hormonal treatments, and social harms, whereas their identities and experiences are also frequently excluded from research.1,16

We recommend that Sexually Transmitted Diseases and other peer reviewed journals adopt policies that require authors to report disaggregated data for trans and nonbinary identities and to use precise and inclusive language. In the absence of available and accurate data on gender identity, we recommend that journals require authors to explicitly acknowledge the limitations of their data. In the sections that follow, we explain how the practice of using binary male/female categories contributes to the erasure of trans and nonbinary experiences, including health inequities, drawing on specific examples within the fields of sexual health and HIV/STIs. We conclude by providing 5 concrete recommendations for inclusive data collection and reporting.

We want to position ourselves in relation to this commentary and our call for researchers and public health practitioners to think critically about how cisnormative assumptions manifest in data collection and reporting. The authors include experts in epidemiology, nursing, infectious disease medicine, social work, and LGBTQ+ activism, all with significant experience working with and within queer and trans communities. Authors also have extensive experience conducting sexual health-related research and public health practice to reduce disparities in HIV/STIs. Authors identify as a gender nonbinary queer person, a genderqueer drag queen, queer trans men, queer cis women, and heterosexual cis women. This commentary aims to reiterate and amplify existing pragmatic, actionable recommendations from trans communities, advocates,
and scientists that improve the inclusion of trans and nonbinary people in research and surveillance.

ERASURE, EXCLUSION, AND MISCLASSIFICATION

Erasure, exclusion, and misclassification are related processes that contribute to informational and health inequities for trans and nonbinary populations. As first described by Bauer et al’s framework, erasure is one of the primary processes through which conscious and unconscious cisnormative bias produces systemic barriers to care for trans communities. In the context of data collection and reporting, erasure is a process that is often enacted through the misclassification and/or exclusion of trans and nonbinary people and identities from surveillance, health systems, and research. It can occur—and be interrupted—at multiple stages of formative research, data collection and analysis, and across institutional levels.

For example, institutions may “lack policies that accommodate trans identities or trans bodies, including the lack of knowledge that such policies are even necessary.”17 This includes medical forms/records, surveillance systems/surveys, and research studies that do not provide trans and nonbinary response options. Institutional erasure also occurs when researchers intentionally or unintentionally exclude trans and nonbinary people from research studies or do not meaningfully engage trans communities, resulting in sample sizes that are too small for statistical inference or from which to draw conclusions. Exclusion constitutes more than the mere absence of trans representation in data; it also occurs when research fails to examine trans experience and, therefore, cannot be generalized to trans and nonbinary populations. In addition, when sample sizes are “too small,” data from trans and nonbinary people are frequently evaluated in aggregate or even aggregated with cisgender populations.18

Erasure is also characterized by a lack of public health policies, guidelines, and recommendations for trans and nonbinary populations, and lack of trans-specific training for health care providers.19,20 For example, the Centers for Disease Control and Prevention does not provide specific recommendations for STI screening, and the US Preventive Services Task Force does not provide specific recommendations for preexposure prophylaxis (PrEP) use among trans and nonbinary people, despite this population having a high incidence of HIV/STIs.21,22 Both sets of recommendations point to a lack of available data on trans populations, for example, stating that: “trials of PrEP enrolled few transgender women and no trials have been conducted among transgender men … its use may be considered in all persons (cisgender and transgender) at high risk of sexual acquisition of HIV.”21

In this way, institutional erasure both produces and is reproduced by informational erasure, or a lack of “knowledge regarding trans people … and the assumption that such knowledge does not exist even when it may.”17 For example, few national surveys or surveillance systems use validated trans-inclusive measures for ascertaining gender.23,24 A review of 71 STI surveillance jurisdictions found that in 2015, 41 (58%) jurisdictions collected data using male/female categories only. In addition, although 11 (15%) jurisdictions collected disaggregated data on trans women and trans men, only 3 reported disaggregated data on trans populations in their STI surveillance reports, suggesting that these data are missclassified or that transgender data were excluded.23 Thus, even when data on trans and nonbinary identities are collected, these data are frequently obfuscated when published/reported.26

In many cases, when data are reported by binary categories (male/female or man/woman), the direction and degree of misclassification is unclear. An assumption is frequently made that the male/female variable available in a data source refers to sex assigned at birth. However, trans and nonbinary people may be misclassified into their sex assigned at birth (eg, trans men may be categorized as female) or be aggregated with cisgender people of the same gender (eg, data on trans women are reported with cisgender women). The direction and degree of misclassification is dependent both on data collection methods and analytic choices made by the researcher. For example, prior studies have shown that when “sex” is ascertained based on the sound of a participant’s voice, such as in the telephone-based Behavioral Risk Factor Surveillance System, one third of trans participants are categorized concordant with the gender (eg, trans men as male and trans women as female) while the remaining two thirds are categorized according to their presumed sex assigned at birth (eg, trans men as female and trans women as male).27,28 In studies that confute sex and gender in their survey questions (eg, “Are you male or female?”), trans respondents may be more likely to provide a response that is concordant with their gender. In studies based on medical or insurance records, male/female categories most likely reflect the current gender marker on an individual’s identification documents, which could either reflect their current gender or their sex assigned at birth.

The aforementioned processes of erasure, exclusion, and misclassification have led to a paucity of high-quality data on the sexual health and overall wellbeing of trans and nonbinary people, because most data on trans health are from small, cross-sectional, and clinical samples.29,30 The confution of sex assigned at birth and gender obscures health disparities among trans and nonbinary populations and neglects important sociostructural factors that impact health outcomes.31,32 In addition, lack of meaningful data on trans experiences impacts how and in what ways funding streams are determined that are aimed at addressing trans health disparities. The absence of data on health outcomes, health care guidelines, and evidence-based treatment and interventions for trans and nonbinary people has contributed to suboptimal or delayed sexual health care, such as lower rates of HIV/STI testing, PrEP uptake, antiretroviral therapy use, or cervical cancer screenings.33–35

RECOMMENDATIONS

We recommend the following practices for trans-inclusive data collection and reporting in alignment with existing guidance.1,9,20,30,36s–39s We believe that peer reviewed journals and editors can play an important role in promoting health equity and the availability of data on trans and nonbinary populations by requiring authors to adopt the following recommendations.

1. Collect data on trans and nonbinary identities using validated and recommended methods.

All studies, health care systems, and surveillance systems should collect data on transgender, nonbinary, and intersex identities using a 2-step question that separately ascertains a respondent’s current gender identity and sex assigned at birth (Fig. 1). The 2-step question has been validated and used in both adult40–42s and adolescent populations43s and is currently recommended by numerous institutional bodies, including the National Academies of Science, the National Institute of Allergy and Infectious Diseases Division of AIDS, Institute of Medicine, Department of Health and Human Services, The Williams Institute, Fenway Health, and The Center of Excellence for Transgender Health at University of California San Francisco.39s,44s We strongly recommend including both nonbinary categories and write-in gender options, since a large proportion of trans people identify with terms beyond the binary categories of man/woman (eg, nonbinary, genderqueer, gender nonconforming, gender fluid, agender, and more). In some settings, it may also be important to include indigenous and additional culturally specific
### Table 1. Examples of How to Discuss Limitations of Existing Data Sources

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical and/or insurance records that collect a binary male/female variable.</td>
<td>“The binary male/female categories available in our data likely reflects the legal gender marker of each patient. This is neither an accurate measure of sex assigned at birth or of gender identity, since many transgender and nonbinary individuals have a legal gender marker that does not reflect their gender identity, and few states allow for a gender-neutral gender marker. We are thus unable to identify transgender and nonbinary patients, who are misclassified in our source data. The direction of this misclassification is also unknown.”</td>
</tr>
<tr>
<td>Survey or interview collects data using imprecise language that conflates sex assigned at birth and gender (eg, “Are you male or female?”).</td>
<td>“The interview script does not distinguish between sex assigned at birth and gender identity, for example, by conflating individuals who are men with male sex assigned at birth. We assume that this measure may more likely reflect a participant’s reported gender identity rather than their sex assigned at birth. Notably, this measure does not allow us to identify transgender and nonbinary participants, who are misclassified in our source data.”</td>
</tr>
<tr>
<td>Data source uses outdated or harmful language.</td>
<td>“We subsequently refer to individuals who selected transgender male to female or who reported female gender identity and male sex assigned at birth as transgender women.”</td>
</tr>
</tbody>
</table>

---

1. What is your current gender identity? (select all that apply)
   - Cisgender Man
   - Cisgender Woman
   - Genderqueer
   - Gender Nonconforming
   - Man
   - Nonbinary
   - Transgender Man/Trans Man
   - Transgender Woman/Trans Woman
   - Two-Spirit¹
   - Woman
   - Additional Category, Please Specify: ________
   - Decline to Answer

2. What was your sex assigned at birth?² (select all that apply)
   - Female
   - Male
   - Intersex³
   - Decline to Answer

The wording for this sample two-step question is adapted from the NIAID’s Division of AIDS Cross-Network Transgender Working Group Recommendations.³⁹

¹ In some settings, it may also be important to include additional indigenous and culturally specific response options, such as third-gender, hijra, fa’afafine, and mahu.

² Emerging literature suggests that some trans and non-binary people oppose the sex assigned at birth question and propose alternatives such as explicitly asking “Do you identify as transgender?” and/or querying anatomy when necessary.⁴⁵–⁴⁷

³ Intersex identity may be best ascertained using a separate question since intersex individuals are often assigned a male or female sex at birth, underscoring the sociocultural construction of binary “sex” categories.⁴⁵

Figure 1. Sample 2-step question.
3. Acknowledge limitations of the data.

We acknowledge that in many cases, researchers are analyzing historical data sources that only collect data on male/female categories; therefore, it is not possible to report data on trans and nonbinary participants. We recommend that authors explicitly state the limitations of their data and their assumptions about the sex assigned at birth and genders of their study population. An implicit assumption is frequently made that a sex/gender variable refers to a participant’s sex assigned at birth. However, as described above, this variable likely misclassifies both sex assigned at birth and gender; and instead, may be a measure of an altogether different construct, such as a participant’s legal gender marker. Examples of how authors can address data limitations are provided in Table 1. In addition, authors should note other local, state, and institutional limitations related to confidentiality concerns or that influence data collection or reporting.

4. Use trans-inclusive language that does not conflate sex assigned at birth and gender.

We recommend that authors use precise and inclusive language that does not conflate sex assigned at birth and gender. Specifically, authors should not generally refer to people assigned female at birth as women and people assigned male at birth as men. In addition, we recommend authors explicitly use the term cisgender when their study population includes cisgender men and cisgender women, rather than using the labels men and women to implicitly mean nontransgender (eg, imprecisely using “men who have sex with men” when exclusively referring to cisgender men who have sex with other cisgender men). Language is always evolving, and there are a number of academic and community-based resources about trans-inclusive language practices and specific terminology to use in place of harmful and/or outdated terms, including language guides developed by the National Institute of Allergy and Infectious Diseases. Some common terms that appear in sexual health research and their alternatives are presented in Table 2. Instead of reproducing terminology that appear in data collection instruments, researchers can acknowledge this limitation in the methodology section (Table 1).

5. Engage with trans and nonbinary communities.

We recommend that researchers elevate the work of trans scientists and health care providers and engage with local transgender-led organizations and their respective communities at all stages of the research process using community-based participatory research or participatory action research approaches. Concepts, definitions, and language used to talk about sex, gender, and transgender bodies and experiences are cultural, historical, and have varied over place and time, and are likely to continue to change. Further, in order to develop and evaluate effective interventions for trans and nonbinary populations, the lived experience of trans people must be reflected in the very design of research by asking questions that are meaningful and relevant to trans lives. Thus, we recommend ongoing practices for including trans and nonbinary people in the process of creating knowledge. In addition, because of the power imbalances common in trans health research and the intersecting institutional harms experienced by these populations, community engagement is also important when conducting secondary analyses of existing data.

Although many of the examples presented in this commentary are specific to HIV/STIs, these recommendations are relevant...
to all fields of behavioral health, public health, and medicine. Trans-inclusive data collection and reporting will facilitate a stronger understanding of the health and wellbeing of trans and nonbinary communities, as well as disparities within trans populations. To optimize the health and policy influence of scientific inquiry, research must account for the impact of social determinants of health, structural racism, and stigma on all LGBTQ+ populations—especially trans and nonbinary people. A fundamental goal of public health is to identify, describe, and explain disparities, with the ultimate aim of eliminating health inequity. However, we will never achieve health equity without changing our institutions and research practices to ensure that trans and nonbinary people are meaningfully and intentionally represented.

REFERENCES

For further references, please see “Supplemental References,” http://links.lww.com/OLQ/A735.