



USAID
FROM THE AMERICAN PEOPLE

Collecting Diverse Data on Gender and Sexuality in Demographic and Health Surveys: An Overview (OPI5)

An Analysis Brief from The DHS Program

Why study how best to collect diverse sexual orientation, gender identity, gender expression, and sex characteristics (SOGIESC) data?

For more than two decades, local demographic and health survey (DHS) steering and technical committees have requested and included diverse sexual orientation, gender identity, gender expression, and sex characteristics (SOGIESC) data to meet their needs. The United States Government, United States Agency for International Development (USAID), and the United Nations have all affirmed that a more robust evidence base is needed to identify and eliminate health disparities based on SOGIESC. Figure 1 provides a timeline of key milestones of collecting SOGIESC data in DHS surveys.

Figure 1. Timeline of Key Milestones Relevant to the Collection of Diverse SOGIESC data in DHS surveys



How should sexual orientation, gender identity, gender expression and sex characteristics (SOGIESC) be measured?

A clear understanding of SOGIESC dimensions is essential to any efforts to collect SOGIESC data. As a USAID-funded project, The DHS Program, as a USAID-funded project, uses terminology from USAID's Automated Directives System.

SOGIESC Dimensions ¹	
Gender	A socially constructed set of rules, responsibilities, entitlements, and behaviors associated with being a man , a woman , or a gender-diverse person , and the relationships between and among people according to these constructs. Though these concepts are linked, the term gender is not interchangeable with the terms women, sex, gender identity, or gender expression.
Gender diverse	Refers to a person with a gender identity beyond the binary categories of man or woman. This can include identifying along the vast diversity of gender identity (e.g., nonbinary , genderqueer , gender nonconforming , genderfluid , etc.).
Gender expression	How a person presents their gender identity outwardly, through acts, dress, behavior, voice, or other perceived characteristics. Gender expression can be described variously as feminine , masculine , both , or neither .
Gender identity	How people perceive themselves and what they call themselves. This can include identifying as a woman or man or as a gender-diverse person along the spectrum of gender identity and gender expression. While gender is a social construct ascribed to individuals, gender identity is self-determined. When someone's sex assigned at birth aligns with their gender identity, the person is cisgender . When someone's sex assigned at birth does not align with their gender identity, the person may identify as a transgender man , transgender woman , nonbinary , or another identity (e.g., gender nonconforming , agender , etc.).
Sex	The designation of a person as male , female , or intersex based on a cluster of anatomical and physiological traits known as sex characteristics. Sex characteristics include external genitalia, secondary sex characteristics (e.g., facial hair, distribution of fat tissue, voice pitch), gonads and internal organs, hormones, and chromosomes. At birth, infants are typically assigned a sex based on visual inspection of external genitalia.
Sexual orientation	The enduring physical, romantic, and/or emotional attraction to other people by sex or gender identity. Common sexual orientations include straight or heterosexual , gay or lesbian , bisexual , pansexual , asexual , queer , and questioning . Sexual orientation is separate from gender identity or gender expression.

¹Definitions from USAID's Automated Directives System Chapter 205

In 2021, the National Institutes of Health asked the National Academies of Sciences, Engineering, and Medicine (NASEM) to recommend measures on sex, gender identity, and sexual orientation for surveys, clinical settings, and administrative forms. NASEM outlines five guiding principles for SOGIESC data collection:

- 1. Inclusiveness** – People deserve to count and be counted
- 2. Precision** – Use precise terminology that reflects the constructs of interest
- 3. Autonomy** – Respect identity and autonomy
- 4. Parsimony** – Collect only necessary data
- 5. Privacy** – Use data in a manner that benefits respondents and respects their privacy and confidentiality

NASEM identifies three dimensions of sexual orientation: behavior, attraction, and identity, and recommends focusing on measuring sexual orientation identity because “it is most consistently tied to material forms of discrimination.” NASEM recommends a single question to capture information on sexual orientation identity.

Q1: Which of the following best represents how you think of yourself?

- Lesbian or gay
- Straight, that is, not gay or lesbian
- Bisexual
- Culturally linguistic-specific term like Two Spirit for American Indian and Alaska Native populations
- I use a different term: [free text]
- (Don't know)
- Prefer not to answer

NASEM recommends collecting data on gender and limiting the collection of data on sex to circumstances where sex traits are relevant. NASEM recommends two direct questions to capture gender identity:

Q1: What sex were you assigned at birth, on your original birth certificate?

- Female
- Male
- (Don't know)
- Prefer not to answer

Q2: What is your current gender?

- Female
- Male
- Transgender
- Culturally linguistic-specific term like Two Spirit for American Indian and Alaska Native populations
- I use a different term: [free text]
- (Don't know)
- Prefer not to answer

Those whose answers to these two questions agree are classified as cisgender while those whose answers differ are classified as transgender.

The NASEM report provides a robust starting point for The DHS Program in considering how best to collect diverse SOGIESC data in DHS surveys. However, there are limitations to applying NASEM's guidance to DHS surveys, including:

The guidance is based primarily on research conducted in the U.S., Canada, and Europe.

The guidance has not been tested in languages other than English and Spanish and DHS surveys are translated into as many as 20 local languages per survey.

NASEM distinguishes between demographic and health surveys and recommends different measures for each. DHS surveys fall into both of these categories.

DHS surveys use proxy reporting where a single household member provides the gender identity information for other household members. Household respondents may not know or accurately report another household member's gender identity.

NASEM's recommended measure for gender identity includes the phrase, “on your original birth certificate” to gather information about sex assigned at birth. This question construction assumes the respondent has a birth certificate which is not the case in many DHS Program countries.

NASEM's recommendation to focus on measuring sexual orientation identity may not be sufficient for DHS surveys which routinely collect information on sexual behavior.

The response options for the question on gender are not mutually exclusive and confuse the concept of sex and gender while reinforcing the notion that cisgender is the norm.

How are SOGIESC data standardly collected in DHS surveys?

The primary aim of DHS surveys is to obtain quantitative data about the health and wellbeing of the population in a given country for use in policy formation, program planning, and monitoring and evaluation. DHS surveys use standard methodology, including standard questionnaires, to be comparable across countries and over time.

DHS Surveys are conducted by a local implementing agency, typically the national statistical office, who convenes the survey's steering and technical committees and determines their membership. The survey steering and technical committees make decisions about the survey scope and content. During questionnaire design, the survey technical committee adapts standard questionnaires and adds country-specific questions to meet their data needs based on the priorities of government ministries, country programs, NGOs, and funding agencies.

Gender-diverse individuals are not excluded from DHS surveys. They are, however, rendered invisible in the data through misclassification, lack of detail, and bias in questions that presume male/female partnerships.

DHS surveys are household-based. Household-based surveys can provide estimates for the general population, but they do not include those living in institutions such as long-term medical or psychiatric care facilities, those who are imprisoned or unhoused, and in most cases do not include individuals living in refugee settlements. Individuals with diverse SOGIESC dimensions are often marginalized and may be more likely to be unhoused or live in an institutional setting, and therefore may be excluded from a household-based survey like a DHS.

DHS surveys use three types of questionnaires – household, individual, and biomarker. The household questionnaire is used to enumerate the individuals living in the household and guests who slept there the prior night. For each household member or guest, interviewers record their name, gender, and current age, as reported by the household respondent. These data are used to identify eligible individuals for the individual interview, typically women age 15–49 and men age 15–54. Table 1 summarizes the SOGIESC-relevant content collected in the DHS standard woman and man's questionnaires.

How the Privacy of DHS Respondents is Protected

Procedures and questionnaires for standard DHS surveys have been reviewed and approved by the ICF Institutional Review Board (IRB). Survey-specific DHS survey protocols are reviewed by the ICF IRB and an IRB in the survey country.

Informed consent is obtained before each interview or biomarker measurement.

Within each household, an eligible respondent may not be interviewed in the presence of another eligible respondent. Results of interviews and biomarker testing are strictly confidential.

Data collection is monitored for adverse events and to ensure data quality. Each data collection team includes a supervisor who is responsible for the safety and well-being of team members and respondents.

Each respondent's interview and biomarker data files are identified only by a series of numbers, including cluster number, household number, and individual number. Data are encrypted and all personally identifiable information is destroyed and geographic coordinates of each survey are displaced at a random distance and in a random direction prior to the public release of datasets.

Table 1. SOGIESC-relevant content in DHS Program woman's and man's questionnaires

Section/Module	Key Indicators	SOGIESC-relevant Notes
Respondent's background	Age, educational attainment, literacy	Proxy reported gender is not confirmed with the respondent
Reproduction	Total fertility rate, children ever born, infant and child mortality rates, menstruation	<p>Woman's Questionnaire: Assumes that a woman has or had a uterus and has a lifetime risk of becoming pregnant, mothers proxy report gender identity for children</p> <p>Man's Questionnaire: Fathers proxy report gender identity for children</p>
Family Planning	Knowledge and current use of family planning methods	Assumes sexual partners are capable of becoming pregnant, and engage in vaginal-penile intercourse
Pregnancy and postnatal care	Antenatal care, assistance during and place of delivery, postnatal care	Not asked of men
Child immunizations	Vaccination coverage	Not asked of men
Child health and nutrition	Treatment of fever and diarrhea, dietary consumption	Not asked of men
Marriage and sexual activity	Marital status and duration, age at first marriage and first sexual intercourse, recent sexual activity	Marriage, cohabitation, and many sexual behavior questions only ask women about male partners and only ask men about female partners. The term 'sexual intercourse' is not defined but is usually understood to mean vaginal-penile intercourse.
Fertility preferences	Desire to limit childbearing, ideal number of children, wanted fertility rates, family planning decision making	<p>Woman's Questionnaire: Assumes a woman has or had a uterus and has a lifetime risk of becoming pregnant</p> <p>Man's Questionnaire: Assumes a man is able to father a child</p>
Employment and gender roles	Employment status, decision-making, land and home ownership, beliefs about wife-beating	Assumes that unions occur between women and men and include questions about the acceptability of a husband beating his wife
HIV	Knowledge of HIV prevention, prior HIV testing, experience of stigma among people living with HIV, prevalence of other STIs	Questions about the experience of genital discharge/sores/ulcers and the acceptability of a wife refusing to have sex with her husband under certain conditions
Other health issues	Prevalence of smoking, alcohol consumption, cervical and breast cancer screening, male circumcision	<p>Woman's Questionnaire: Assumes a woman has or had a uterus, cervix, and breasts. Cervical and breast cancer screening questions are not asked of men.</p> <p>Man's Questionnaire: Assumes a man has a penis. Penile circumcision is not asked of women.</p>
Domestic violence module	Experience of physical, sexual, and emotional violence	Intimate partner violence is defined as occurring between a woman and a man only
Female genital cutting (FGC) module	Prevalence of FGC, beliefs about FGC	Assumes a woman has or had a vulva, clitoris, and labia. Mothers proxy report information on their daughters. Experience of FGC is not asked of men.
Fistula module	Knowledge and prevalence of fistula	Assumes that a woman has a vagina. Experience of fistula is not asked of men.

How have DHS surveys collected more diverse data on sexual orientation, gender identity, gender expression and sex characteristics?

For more than two decades, DHS surveys have included survey-specific questions to collect more diverse SOGIESC data.

India

The census in India first expanded the gender options in 2011 with an ‘other’ category and in 2014, the Supreme Court of India legally recognized a third gender. The Government of India requested a change in the questionnaire in the 2015–16 India DHS, but the request came after the questionnaire was finalized and data collection was about to begin. The 2019–21 India DHS added a third gender option, transgender, in the household roster. There was no individual questionnaire administered to those identified as transgender. Transgender individuals were also not eligible for biomarker measurement or testing. Data collectors relied solely on the household respondent’s responses to the question “Is (NAME) male or female or transgender?” Transgender individuals were shown separately from females and male individuals in only one table in the final report: prevalence of any disability among the household population age 15 and older.

The 2019-21 India DHS household roster asked
“Is (NAME) male or female or transgender?”

The household questionnaire respondent provided this information for each household member.

South Africa

The South African parliament legalized same-gender marriages in 2006. The 2016 South Africa DHS included questions on the gender of individual women’s and men’s partners. Instead of asking the standard question which specifies the opposite gender when asking about marriage or living together as if married, the question for both women and men was “Are you currently married or living together with someone as if married?” This was then followed by the question “Is this person a woman or a man?”

The 2016 South Africa DHS asked

“Are you currently married or living together with someone as if married?”

If the respondent said yes, they were then asked “Is this person a woman or a man?”

No same-gender specific additions were made to any tables. Those in same-gender partnerships were included in the standard category of ‘married or living together as if married’. A few tables in the final report were modified to note that individuals in same-gender relationships were not asked questions on women’s empowerment nor polygyny. Additionally, only women in a relationship with a man were eligible for the domestic violence module. The reasoning behind this decision was that these questions were developed in a framework of different gender relationships. A slightly different question was included in the 2003 South Africa DHS – “Are you currently married or living with a partner?” with options for ‘yes, married/living with a woman’ and ‘yes, married/living with a man’. The 2003 DHS data are not publicly available, but a final report is published. There were no tabulations of same-gender marriages.

Colombia

The 2015 Colombia DHS included the most survey-specific diverse SOGIESC additions. This was the first national-level survey in Latin America to include these topics. In Colombia, transgender individuals over the age of 18 were granted the right to change their legal gender in 2015 and same-gender marriage was legalized in 2016.

The household roster included four options for the gender of each person – ‘male’, ‘female’, ‘trans male’, and ‘trans female’. The first question in the individual interview then asked the respondent to confirm their gender – “Since we are surveying both men and women, I need to ask you the following question to determine the type of questionnaire to administer. Are you a man or a woman or a transgender woman or a transgender man?” There were still just two individual questionnaire types, one for women and transgender women and one for men and transgender men.

The first question in the 2015 Colombia DHS individual questionnaire was

“Since we are surveying both men and women, I need to ask you the following question to determine the type of questionnaire to administer. Are you a man or a woman or a transgender woman or a transgender man?”

The individual questionnaires included a direct question on sexual orientation with response options of ‘heterosexual’, ‘homosexual’, or ‘bisexual’. A full marriage history was obtained and for each relationship, the gender of the partner was asked. Questions on recent sexual activity assumed heterosexual relationships in terms of the gendered terms used in Spanish.

Finally, all respondents were asked about their opinions on the rights of LGBT individuals in Colombia. The final report included a table on the percent distribution of respondents by sexual orientation, percent distribution of the gender of respondents’ partners, and a table about attitudes toward the LGBT population. No other tables split out sexual orientation, likely due to the very small sample size.

Other Surveys

Table 2 presents an overview of all DHS surveys that included diverse SOGIESC topics including India, South Africa, and Colombia. The surveys in Table 2 only included same-gender sexual behavior questions among men. The questions included ever having had sex with a man and recent sex with a man. In Myanmar and the Philippines, DHS surveys also asked men about condom use the last time they had sex with a man, added to improve HIV prevention programs. Cambodia was the only country to include these questions in multiple surveys.

Table 2. DHS Surveys that include diverse SOGIESC-related questions

Survey	Diverse SOGIESC topics
2002 Dominican Republic DHS	Same-gender sexual behavior
2003 Philippines DHS	Same-gender sexual behavior
2003 South Africa DHS	Gender of spouse/partner
2004 Bangladesh DHS	Same-gender sexual behavior
2005 Cambodia DHS	Same-gender sexual behavior
2010 Cambodia DHS	Same-gender sexual behavior
2014 Cambodia DHS	Same-gender sexual behavior
2015 Colombia DHS	Transgender gender option, sexual orientation, same-gender sexual behavior, attitudes toward LGBT persons
2015–16 Myanmar DHS	Same-gender sexual behavior
2016 South Africa DHS	Gender of spouse/partner
2019–21 India DHS	Transgender gender option
2021 Cambodia DHS	Same-gender sexual behavior

What are some key considerations for the collection of SOGIESC data in DHS surveys?

Autonomy

The request to collect diverse SOGIESC data in a DHS survey must originate from the survey steering or technical committee. Such requests should be reviewed by groups working with and inclusive of local LGBTQI+ communities to ensure the data collected represent the authentic, culturally specific experiences of local LGBTQI+ people. Local groups can help ensure SOGIESC data needs are identified (e.g., prevalence of health disparities by gender identity or sexual orientation, prevalence of gender-based violence among LGBTQI+ communities, etc.). The survey can then be tailored to meet those needs, including estimating the needed sample size and anticipating response rates. Failure to adequately plan for sample size and responses rates could result in underestimation or erasure of LGBTQI+ people's experiences.

Legality

An overarching concern in the collection of diverse SOGIESC data is the legality of same-gender sexual relationships and transgender and gender-diverse identities in the country conducting a DHS survey. There is universal agreement that the safety and security of respondents is the top priority. Collecting information on illegal behaviors is not insurmountable, as other DHS survey topics, such as child marriage and female genital cutting, share this challenge.

Interviewer Training and Respondent Safety

Both the survey protocol and data collection team training need to address interviewer and respondent safety and how to respond to community or individual resistance to questions about sexuality and gender that may be outside of commonly accepted social norms. Training materials for diverse SOGIESC-specific questions are critical for their inclusion in a DHS survey, as data collection team members need to be sensitized to these topics to ensure that respondents understand what is being asked. While DHS policies already exist for respondent safety and privacy, there may be additional considerations required when identifying possibly marginalized individuals.

Precise Terminology

Terminology concerning gender identity and sexual orientation is contextually specific. The NASEM recommended response options for gender have received valid criticism for a lack of specificity and for not being mutually exhaustive categories, e.g., a transgender woman is a woman. Qualitative research may be necessary prior to fielding a DHS survey that intends to include questions on gender diversity. Local LGBTQI+ organizations should be consulted during questionnaire design and could even be involved in the pre-test and training material adaptation.

Inclusiveness

Simply changing the pronouns used or rewording questions to be gender neutral is not sufficient to ensure robust inclusion of people with diverse SOGIESC dimensions in DHS surveys. As detailed above in Table 1, the DHS woman's and man's questionnaires are intentionally different and ask about issues deeply rooted in biology presumed to correspond to a binary gender assigned at birth. The 2015 Colombia DHS provides one example of how to overcome proxy reporting of gender identity by asking respondents to confirm their gender before the interview, but room for improvement remains.

To ensure meaningful inclusion and representation of LGBTQI+ individuals in DHS surveys, survey steering and technical committees, along with USAID and The DHS Program staff, must consider and contend with competing data needs and priorities while managing the complexities of adapting survey instruments originally designed to measure fertility to be inclusive of individuals with diverse sexual orientations, gender identities, gender expression, and sex characteristics.

This brief summarizes The DHS Program's Occasional Paper 15 by Sarah M. Balian, Blake Zachary, and Kerry L.D. MacQuarrie with funding from The United States Agency for International Development through The DHS Program implemented by ICF. The full report is available at: <https://www.dhsprogram.com/publications/publication-op15-occasional-papers.cfm>