



DHS Survey Design: Collecting Diverse Gender and Sexuality Data

This brief describes considerations for the collection of diverse data on sexual orientation, gender identity, gender expression, and/or sex characteristics (SOGIESC) in Demographic and Health Surveys (DHS). The information is based on existing research, technical guidance, and The DHS Program's experience in assisting countries to collect diverse SOGIESC data in nationally representative household-based surveys.

Why collect diverse SOGIESC data in a DHS survey?

For more than two decades, local steering and technical committees have requested and included diverse SOGIESC data in DHS surveys to meet their needs. The United States Government, the 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.

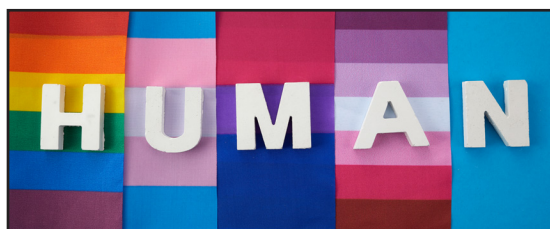
Collecting diverse SOGIESC data in a DHS survey can help bring visibility to the lesbian, gay, bisexual, transgender, and queer (LGBTQ) population, as well as provide information on topics such as education, employment, access to information, marriage and sexual activity, reproduction and family planning, nutrition, HIV, malaria, domestic violence, alcohol and tobacco use, and other health issues to guide programs and policies to promote health and well-being.

Can diverse SOGIESC data safely be collected in DHS surveys?

An overarching concern in the collection of SOGIESC data is the legality of same-gender sexual relationships and gender in all its diverse identities (e.g. non-binary, transgender, third gender individuals, etc.) which varies dramatically across countries that conduct DHS surveys. Even if something is not illegal, there can still be tremendous shame and stigma surrounding it. The safety and security of DHS survey respondents and data collection teams are of paramount importance.

The request to collect diverse SOGIESC data in a DHS survey must originate from the survey's steering or technical committee, indicating local desire and commitment to collect and use these data. Consulting with local LGBTQ-serving organizations is important to ensure the data collected are relevant and that survey protocols, training materials, and resources for respondents and data collection teams appropriately assess risk and follow the principle to 'do no harm'.

DHS surveys have strong standard privacy protections including but not limited to: Institutional Review Board review in both the US and the survey country, obtaining informed consent before each interview, conducting interviews in private settings, monitoring data collection for adverse events, data collection supervisors who are responsible for the safety of respondents and data collection team members, encryption of all data during data collection, the destruction of all personally identifiable information before data are distributed, and the use of serial numbers to identify respondents in data files.



Which SOGIESC-related indicators can be collected in a DHS survey?

Indicator	Definition	Denominator/Universe
Gender Identity*		
Cisgender women age 15-49	Percentage of individuals who currently identify as a woman and were assigned female at birth	Individuals age 15-49
Cisgender women all ages	Percentage of de jure population identified as a woman by the household interview respondent	De jure household population
Cisgender men age 15-49	Percentage of individuals who currently identify as a man and were assigned male at birth	Individuals age 15-49
Cisgender men all ages	Percentage of de jure population identified as a man by the household interview respondent	De jure household population
Transgender women age 15-49	Percentage of individuals who currently identify as a woman who were assigned another sex at birth	Individuals age 15-49
Transgender women all ages	Percentage of de jure population identified as a trans woman by the household interview respondent	De jure household population
Transgender men age 15-49	Percentage of individuals who currently identify as a man who were assigned another sex at birth	Individuals age 15-49
Transgender men all ages	Percentage of de jure population identified as a trans man by the household interview respondent	De jure household population
Gender-diverse individuals age 15-49	Percentage of individuals who do not currently identify as a woman or a man	Individuals age 15-49
Gender-diverse individuals all ages	Percentage of de jure population identified as a gender other than woman or man by the household interview respondent	De jure household population
Sexual Identity*		
Lesbian individuals age 15-49	Percentage of individuals who currently identify as lesbian	Individuals age 15-49
Gay individuals age 15-49	Percentage of individuals who currently identify as gay	
Bisexual individuals age 15-49	Percentage of individuals who currently identify as bisexual	
Heterosexual individuals age 15-49	Percentage of individuals who currently identify as heterosexual	
Individuals age 15-49 with another sexual orientation	Percentage of individuals who currently identify as another sexual orientation	
Individuals age 15-49 with unknown sexual orientation	Percentage of individuals who currently don't know their sexual orientation or refuse to answer	
Composite Gender Identity and Sexual Identity		
LGBTQ individuals age 15-49	Percentage of individuals who currently identify as lesbian, gay, bisexual, trans, gender-diverse, queer	Individuals age 15-49
Non-LGBTQ individuals age 15-49	Percentage of individuals who currently identify as heterosexual, cisgender woman, cisgender man	
Same-Gender Partnership		
Same-gender partnership among women age 15-49	Percentage of women who report their current partner is a woman	Women age 15-49
Same-gender partnership among men age 15-49	Percentage of men who report their current partner is a man	Men age 15-49
Same-Gender Sexual Behavior		
Same-gender sexual intercourse among women age 15-49	Percentage of women who have had sexual intercourse with a woman in the past 12 months	Women age 15-49
Same-gender sexual intercourse among men age 15-49	Percentage of men who have had sexual intercourse with a man in the past 12 months	Men age 15-49
Attitudes Toward LGBTQ Individuals		
Favorable or discriminatory attitudes toward LGBTQ individuals	Percentage of individuals who have favorable/discriminatory attitudes toward LGBTQ individuals	Individuals age 15-49

*Gender identity and sexual identity indicators are illustrative. Response categories and indicator definitions will vary based on the country-specific context.

How can DHS surveys accurately capture information on gender and sexuality which are culturally specific concepts?

Working with a local LGBTQ-serving organization from the beginning is essential to ensure the DHS survey protocol, instruments, and data represent the authentic, culturally-specific experiences of LGBTQ individuals in the survey country.

DHS surveys are conducted in between 2 and 20 languages, depending on the linguistic diversity in the country. Translation must be consistent since some languages do not have separate words for sex and gender, so these constructs must be clearly defined. Sexual orientation must also be clearly defined, as it may include identity, attraction, and/or behavior. Cognitive testing may be needed to verify respondent comprehension.

How can diverse SOGIESC data contribute to achieving the Sustainable Development Goals (SDGs)?

The SDGs include “a pledge that no one will be left behind” and a focus on reaching the most marginalized. DHS surveys can provide estimates for 50 indicators across 10 of the 17 SDGs.

DHS surveys use a household questionnaire to enumerate all individuals living in the household and guests who slept there the prior night. A single respondent provides age, gender, and school attendance information for all these individuals (proxy reporting). Though self-reported gender identity is the most reliable, proxy reporting of gender identity has been shown to be reasonably reliable. In contrast, sexual orientation information cannot be reliably obtained via proxy reporting.



Considerations for Inclusion

Validity:

Diverse SOGIESC data have been collected in censuses and nationally representative surveys in South and North America, Europe, Africa, Asia, and Oceania, including 12 DHS surveys. Validity depends on ensuring the survey instruments reflect the culturally-specific understanding of gender and sexuality and the diversity of identities in the survey country. Questions and response options must be understood by both LGBTQ and non-LGBTQ individuals, which may require cognitive testing of the survey instruments. Collaborating with local LGBTQ-serving organizations can help determine which indicators are most relevant to collect, what is the appropriate wording for survey instruments, and what changes to make based on the pre-test results.

Impact on planning:

Ensuring robust LGBTQ inclusion in DHS surveys must be adequately planned for during survey design. A local LGBTQ-serving organization to support the survey should be identified as early as possible. Questionnaire design, adaptation, and translation will require additional time to develop appropriate instruments based on the content and scope of the DHS survey.

Impact on quality:

Data collection teams must be sufficiently trained to collect high-quality SOGIESC data. A local LGBTQ-serving organization can help develop training sessions, manuals, and facilitate the training for data collection team members. Pre-testing is also crucial for quality data.

Impact on cost:

Collecting diverse SOGIESC data in a DHS survey will increase the total cost, as developing and translating survey instruments will require additional effort, pre-test and training may take longer, cognitive testing may be needed, and the local LGBTQ-serving organization should be compensated for their contributions to the survey.

DHS Survey Design:

Frequently Asked Questions about collecting diverse gender and sexuality data

The DHS household questionnaire gathers information about housing conditions to estimate SDG indicators on population access to or experience of:



- Electricity
- Safe and readily available water
- Sanitation facilities
- Hand-washing facilities
- Clean fuels and technology for cooking, heating, and lighting
- Food insecurity

To identify possible disparities, DHS survey estimates for these SDG population-level indicators could be disaggregated by proxy-reported gender identity, but not by sexual identity.

The DHS individual questionnaires cover a wide range of topics. Individuals age 15-49 self-report information about themselves which can be used to estimate SDG indicators on:



- Child marriage
- Adolescent birth rates
- Intimate partner violence
- Female genital mutilation
- Use of banks and/or mobile money providers
- Use of family planning
- Informed decision-making for sexual and reproductive health
- Tobacco use
- Mobile phone ownership and internet use

DHS survey estimates for these SDG individual-level indicators could be disaggregated by self-reported gender identity, sexual identity, gender of partner, sexual behavior, or a composite indicator of gender and sexual identity.

What are the key sample size considerations for estimating SOGIESC indicators at the national level?

It can be challenging to collect diverse SOGIESC data due to the size of the LGBTQ population and the stigma and discrimination associated with being LGBTQ. However, there are often other studies that have collected information about key populations in the survey country which can be used as a baseline to estimate the percentages of LGBTQ individuals in the sample. A review of global nationally representative surveys conducted since 2019 shows that usually more than 90% of respondents identify as heterosexual. Non-response varies from 1% to 4%. Regarding gender identity, between 0.1% and 0.9% of the population identify as gender-diverse.

Indicator	Expected LGBTQ disclosure	Sample size for minimum number of cases	Sample size for between group analyses
In countries with low levels of tolerance and low levels of LGBTQ rights			
Gender-diverse identity	0.1%	50,000	400,000
Sexual orientation or composite LGBTQ indicator	<2%	2,500	20,000
In countries with low levels of tolerance and medium levels of LGBTQ rights			
Sexual orientation or composite LGBTQ indicator	2% – 4%	1,429 – 2,500	11,429 – 20,000
In countries with medium levels of tolerance and medium levels of LGBTQ rights			
Gender-diverse identity	0.5%	10,000	80,000
Sexual orientation or composite LGBTQ indicator	≥4%	1,250	10,000