2025 ANNUAL LGBTQ+ COMMUNITY SURVEY

Methodology and Data Quality Report

AUTHORS

Charleigh Flohr, Associate Director, Research, Human Rights Campaign Foundation
Tari Hanneman, Senior Director, Research & Education, Human Rights Campaign Foundation

INQUIRIES

Press inquiries: Press@HRC.org
Research inquiries: Research@HRC.org

1. OVERVIEW

The Human Rights Campaign Foundation's 2025 Annual LGBTQ+ Community Survey (ALCS) was designed to generate robust data on the experiences, attitudes, and well-being of adults in the United States, with specific attention to LGBTQ+ populations.

The 2025 survey features two parallel, nonprobability samples. It includes one for sexual and gender minority (SGM) adults and one for non-sexual and gender minority adults. The parallel samples enable both within-group and comparative analyses.

The following report documents how the raw survey data were cleaned, validated, and weighted to produce analytic datasets used in HRC Foundation research and analysis publications. The sections below describe data cleaning and quality control, weighting procedures, scaling for national representation, validation, and other quality considerations.

2. SURVEY QUESTIONNAIRE & SAMPLE DESIGN

Population and Fieldwork

The analytic universe includes adults aged 18 and older residing in the United States. Data collection occurred between September 29, 2025 and October 27, 2025. Parallel questionnaires were administered online. The LGBTQ+ sample was gathered in Qualtrics using HRC Foundation's panel directory of LGBTQ+ adults in the United States. In partnership with HRC, PSB Insights fielded a survey with n=5,003 non-LGBTQ+ identifying Gen Pop consumers from September 29th through October 27th. To ensure the survey sample reflected the demographic distribution of the general population in the US, PSB set representative quotas for respondents' race, age, geography, education level, and gender.

All respondents completed identical instruments on core-questionnaires containing standardized measures of demographics, attitudes, and experiences to permit direct comparison between LGBTQ+ and non-LGBTQ+ adults. In some rare questions, survey questions were only asked to LGBTQ+ or non-LGBTQ+ adults or subsets of the population such as private/public sector workers in large companies. This was because the survey question(s) would only be applicable to that specific subpopulation.

Sampling Goals

Two separate sample goals were set and met:

- SGM Sample (N = 10,000): Adults identifying as lesbian, gay, bisexual, transgender, queer, or another minority sexual or gender identity.
- Non-SGM Sample (N = 5,000): Adults identifying as heterosexual and cisgender.

Each sample was recruited independently to avoid contamination or overrepresentation of one group. Quotas for race/ethnicity, age, and educational attainment were monitored during fielding to ensure broad coverage.

3. DATA CLEANING & CASE CONTROL

Initial Dataset

The initial combined dataset contained more than 15,000 responses across both samples. Quality control was conducted prior to merging.

Quality Checks

A multi-step cleaning protocol was implemented using Stata:

1. **Completion Rate**: Cases completing the survey in fewer than four minutes were removed, as were partial completes under 35% progress.
2. **Duplicate Detection**: Potential duplicates were flagged using combinations of IP address, browser fingerprint, and respondent email. Stata bysort and duplicates commands as well as conditional logic identified instances where multiple records shared identical identifiers.
3. **Retention Rules**: When duplicates were detected, only the record with 100% completion and the highest data quality score was retained.
4. **Straightlining**: Responses showing invariant patterns (90% straightline) across matrix items were removed.

Final Analytic Dataset

After cleaning, approximately 9,600 valid LGBTQ+ respondents and 5,000 valid non-LGBTQ+ respondents remained. These serve as the analytic base for all HRC 2025 products.

Weight Construction and Calibration

The dataset required weighting to correct for error resulting from non-probability sampling and to align demographic distributions with external population benchmarks. The weighting strategy employed a two-stage process:

1. Iterative Proportional Fitting (IPF or "Raking") within each parallel sample, and
2. Scaling of the SGM sample weights to ensure the combined dataset reflects the national share of LGBTQ+ adults.

Two weight variables were generated:

- *allwt* — within-sample raked weight;
- *allwt_scaled* — scaled weight used for national population estimates.

Within-Sample Iterative Proportional Fitting

Each parallel sample was weighted independently to reflect its respective reference population. Data for SGM distributions come from the Gallup Poll Social Series and Williams Institute. In 2025, Gallup reported that 9.3% of U.S. self-identified as LGBTQ+ people (Jones, 2025). In a 2023 report, the William's Institute pooled 2020-2021 data from the Behavioral Risk Factor Surveillance System and provides age and region distributions (Flores & Conron, 2023). Proportions for race and educational attainment come from the William's Institute's LGBT Data and Demographics project, which used data from the Gallup Daily Tracker (Kastanis et al., 2019). Census Bureau data are used to provide the official benchmark for the size and composition of the U.S. adult population.

Each parallel sample (SGM and non-SGM) was calibrated separately using iterative proportional fitting (IPF) implemented through the Stata command ipfweight. This procedure adjusts initial weights so that the weighted sample marginal distributions match a set of external population targets across multiple categorical variables simultaneously.

Let:

$$w_i^{(0)} = 1$$
$$M_j = level\ m\ in\ variable\ j$$
$$T_{jm} = population\ parameter\ of\ M_j$$
$$E_{jm}^{(t)} = observed\ proportion\ of\ M_j$$
$$I_{jm} = individual\ index\ of\ M_j$$

At each iteration, $t + 1$, weight w_i for each respondent i is updated according to:

$$w_i^{t+1} = w_i^t \prod_{j=1}^{J} \frac{T_{j,mi}}{E_{j,mi}^t}$$

Decomposition for Non-SGM Benchmarks

Because direct benchmark data for non-LGBTQ+ adults are limited, HRC Foundation constructed implied benchmarks using a decomposition approach and population parameters.

$$pT(c) = s\big(pSGM(c)\big) + (1 - s)\big(pCisHet(c)\big)$$

Where:

$$pT(c) = overall\ adult\ population\ in\ category\ c$$
$$pSGM(c) = SGM\ population\ in\ category\ c$$
$$pCisHet(c) = nonSGM\ population\ in\ category\ c$$
$$s = overall\ share\ of\ SGM\ adults$$

Solving for the non-SGM proportion:

$$pCisHet(c) = \frac{pT(c) - s\big(pSGM(c)\big)}{1 - s}$$

Scaling for National Representation

Following within-sample raking, both datasets were merged for national-level analysis. To ensure the combined sample matched the national proportion of SGM adults, weights for the LGBTQ+ sample were scaled relative to the non-LGBTQ+ sample.

Let:

$$\sum w_{sgm} = total\ unscaled\ SGM\ weights$$
$$\sum w_{non} = total\ unscaled\ nonSGM\ weights$$
$$s = SGM\ population\ parameter\ of\ 0.093$$

The scaling factor is represented by:

$$k_{sgm} = \frac{s(\sum w_{non})}{(1-s)(\sum w_{sgm})} \text{ with computed scaled weights} = w_{SGM,scaled} = \big(k_{sgm}\big)\big(w_{sgm}\big)$$

This adjustment ensures that SGM adults comprise 9.3% of the weighted national dataset. All cases retain their raked weights as *allwt* with the scaling factor applied in *allwt_scaled* to compute national estimate (full sample), while *allwt* allows for within-group analyses (LGBTQ+ vs. Non-LGBTQ+ parallel samples).

Variance Estimation and Design-Based Inference

All analyses using weighted data use a design-based inference framework. The survey design is declared in Stata:

```
svyset UniqueID [pweight = allwt_scaled], vce(  )
```

When analyzing each subsample separately (SGM or non-SGM), the unscaled weights (allwt) were used; when combining samples to generate national estimates or LGBTQ+/non-LGBTQ+ comparisons, the scaled weights (allwt_scaled) were applied.

4. REFERENCES

Flores, A. R. & Conron, K. J. (2023). *LGBT Adults in the U.S. Population.* The Williams Institute.

Gallup (2025). *LGBTQ+ Identification Rises to 9.3%.*

Kastanis, A., Conron, K., & Gates, G. J. (2019). *LGBT Data & Demographics Project.* The Williams Institute.

U.S. Census Bureau (2020). *American Community Survey 1-Year Estimates.*

HRC Research & Education Program (2025). *Iterative Proportional Fitting Specifications for Parallel SGM/Non-SGM Analysis.*