Patterns of Non-Response to SOGI Questions: Findings from a Brazilian Survey

Mariel Gouvea Gruppi[1]

Cássio M. Turra[1]

Simone Wajnman[1]

[1]Center for Development and Regional Planning (Cedeplar), Federal University of Minas Gerais, Belo Horizonte, Brazil.

Abstract

As questions on sexual orientation and gender identity (SOGI) are increasingly included in household surveys, understanding patterns of nonresponse to these items has become crucial—particularly in contexts marked by stigma, inequality, and limited institutional support. This study investigates nonresponse to SOGI questions using two Brazilian datasets: the 2019 National Health Survey (PNS), which employs self-reporting, and the 2021 District Household Sample Survey (PDAD), which uses proxy reporting. We examine how sociodemographic characteristics and refusal to answer other sensitive items predict nonresponse, using descriptive statistics and logistic regression models. Our findings reveal that nonresponse rates vary significantly by survey design: while SOGI nonresponse is relatively high in the self-reported PNS, it is markedly lower in the proxy-based PDAD—raising questions about data completeness versus disclosure accuracy. Results also show that nonresponse is more likely among men, unmarried individuals, those with lower education, and respondents who refuse to answer other personal questions. These patterns underscore the role of cognitive, emotional, and contextual barriers in shaping SOGI disclosure. This study contributes to emerging debates in queer demography and survey methodology, offering insights for improving data quality and advancing inclusive statistical practices in Brazil and beyond.

Introduction

Collecting information on sexual orientation is a relatively recent development in several countries, particularly in Brazil. It was only in 2022 that data from the 2019 National Health Survey (NHS) became available, marking the first instance of a national survey inquiring about the sexual orientation of Brazilians as part of its diverse questionnaire. It represents the inaugural nationally representative study to collect self-declared data on sexual orientation from Brazilian people.

Regarding gender identity in Brazil, only the 2021 PDAD has thus far collected information on transgender and non-binary individuals in its sample, alongside including the categorization of sexual orientation. The latter survey was a household survey conducted in one of Brazil's regions, the Federal District, home to the country's capital, Brasília.

The importance of identifying sexual and gender minorities in household surveys is justified by various findings that indicate poorer mental health conditions for gay and lesbian individuals (Carpenter, Eppink, Gonzales, & McKay, 2021), disadvantages in the labor market for lesbian, gay and bisexual individuals (Drydakis, 2022; Suliano, Filho, & Irffi, 2021), as well as a higher risk of suicide and experiencing violence among sexual and gender minorities (LeVasseur, Kelvin, & Grosskopf, 2013; Pinto et al., 2020). Additionally, a Gallup study on the United States suggests that 20% of Generation Z, those born between 1997 and 2003, and only 2.6% of Baby boomers, those born between 1946 and 1964, identify as a sexual or gender minority, which suggests that this is a population with a large growth in the coming years. However, there is much discussion about including a question that aims to collect sensitive information in a questionnaire, because it

could affect the whole response rate. As it is the case of sexual orientation and gender identity questions, due to its private and stigmatized nature, not everyone would be willing to disclose their identity to someone else, which could affect the survey response rate, thereby impacting the quality of the data (Young & Rater, 2021).

Objectives and Focus

Given the challenges outlined above, this study seeks to advance the emerging body of research on nonresponse to sexual orientation and gender identity (SOGI) questions in population-based surveys. While international studies have begun to explore patterns and predictors of nonresponse—particularly in high-income countries—there remains a critical gap in Latin America, and Brazil in particular, where research on SOGI disclosure and refusal is still incipient. This study aims to fill part of this gap by conducting a comparative analysis of two distinct Brazilian surveys: the 2019 National Health Survey (PNS) and the 2021 District Household Sample Survey (PDAD). The PNS represents the first nationally representative effort to collect self-declared sexual orientation data, while PDAD is the first survey in Brazil to include both sexual orientation and gender identity questions—albeit using a proxy response method.

This research is innovative for at least two reasons. First, it is the first known Brazilian study to jointly analyze nonresponse to both sexual orientation and gender identity items using two complementary datasets. Second, it brings attention to how different survey designs and response modalities (self-reporting vs. proxy reporting) interact with sociodemographic characteristics and refusal behavior to shape the visibility—or erasure—of LGBTQ+ populations in national statistics.

The study pursues three main objectives:

1. To describe patterns of nonresponse to SOGI questions in both surveys;
2. To estimate the likelihood of nonresponse using binomial logistic regression models, with a focus on calculating Odds Ratios;
3. To test specific hypotheses regarding the determinants of nonresponse, including:
    - the effect of survey methodology (self vs. proxy reporting) on response rates and respondent profiles;
    - the role of refusal to other sensitive questions (e.g., age of first sexual intercourse, income) as a predictor of SOGI nonresponse;
    - and the influence of social position variables, such as race, education, and marital status, on the likelihood of disclosure.

While the PDAD survey is geographically limited to the Federal District, it offers a unique opportunity to assess the feasibility and effectiveness of including SOGI items in local-level data collection. Together, the two surveys provide complementary insights into how identity disclosure operates in different institutional and methodological contexts. By addressing both the structural and individual-level drivers of nonresponse, this study contributes to improving data quality, refining survey instruments, and ultimately expanding the representativeness of LGBTQ+ individuals in demographic research.

Data and Methods

This study is based on data from two large-scale Brazilian household surveys that included questions on sexual orientation and/or gender identity: the 2019 National Health Survey (Pesquisa Nacional de Saúde – PNS) and the 2021 District Household Sample Survey (Pesquisa Distrital por Amostra de Domicílios – PDAD). These surveys differ markedly in scope, methodology, and method of data collection, offering a unique opportunity to examine how these factors influence nonresponse patterns to SOGI questions.

The PNS 2019 is a nationally representative health survey conducted by the Brazilian Institute of Geography and Statistics (IBGE) in partnership with the Ministry of Health. It included, for the first time, a direct question on sexual orientation in its "Sexual Activity" module, answered only by one randomly selected adult per household. PNS employed self-response methods, offering closed response categories and privacy-enhancing protocols (e.g., private entry on interviewer's device). It did not include a gender identity question, nor did it distinguish between sex and gender in the questionnaire. Respondents could select "don't know" or "refused to answer," and these were recoded as instances of nonresponse for this study.

The PDAD 2021, in contrast, is a regional household survey conducted in the Federal District by IPEDF (formerly CODEPLAN). It is the only Brazilian survey to date to include both sexual orientation and gender identity questions, and it did so using a proxy response format: one person in the household provided answers on behalf of all other adult residents. The questions were open-ended and designed to capture a wider spectrum of identities. For sexual orientation, respondents could indicate either identity labels (e.g., gay, lesbian, bisexual) or patterns of attraction (e.g., "attracted to men/women/both"). For gender identity, the survey used a two-step question: first asking about the sex assigned at birth, then asking with which gender the person currently identifies. As in the PNS, respondents could also answer "don't know" or "refused," which constitute the nonresponse category.

For both surveys, the unit of analysis is the individual aged 18 years or older. The dependent variables in the analysis are two binary indicators of nonresponse—one for sexual orientation (present in both surveys) and another for gender identity (present only in PDAD). Individuals who selected "don't know" or "refused to answer" were coded as 1 (nonresponse); those who provided a valid identity were coded as 0.

The independent variables include key sociodemographic characteristics available in both surveys as age group, race/skin color (aggregated into "White/Yellow" and "Black/Brown/Indigenous"), education level, and marital status. Additional covariates are included where available. In the PNS, we control for variables such as respondent type (self vs. proxy), co-residence with mother, and region type (capital, metropolitan, or other). We also include behavioral indicators, such as refusal to answer other sensitive questions (e.g., age at first sexual intercourse, sexual activity in the last 12 months). In the PDAD, the only sensitive variable with a refusal option is monthly income, which is also included as a predictor.

To account for sample design and stratification, all analyses use survey weights. The binary logistic regression models are estimated for each outcome of interest (nonresponse to sexual orientation and nonresponse to gender identity). Odds Ratios (ORs) are computed to assess the relative likelihood of refusal across categories of explanatory variables.

Results

The comparative analysis of nonresponse rates to SOGI questions reveals important differences across surveys, survey modes, and question types. As shown in Table 1, the overall nonresponse rate to the sexual orientation question in the nationally representative PNS (2019) was 3.35%, whereas the rate in the PDAD (2021), restricted to the Federal District, was significantly lower at 0.64%. When isolating the PNS data for the Federal District, the nonresponse rate increased to 4.90%, further emphasizing the discrepancy between the two surveys. For gender identity, available only in the PDAD, the nonresponse rate was remarkably low—just 0.17%.

These findings suggest that survey design and data collection method play a critical role in shaping disclosure patterns. While the PNS uses self-reported data and asks sensitive questions directly to individuals, the PDAD adopts a proxy response approach, where one household member reports on behalf of all others. Proxy reporting appears to significantly reduce item nonresponse, though it may also suppress disclosure of minority identities due to third-party filtering or lack of knowledge. Additionally, the nonresponse to gender identity is lower than that of sexual orientation in the PDAD, possibly reflecting differences in perceived sensitivity or social norms surrounding gender versus sexuality—particularly in proxy settings.

Table 1: Nonresponse rate for SOGI question, by survey, Brazil and DF

	Brazil	DF	
	PNS	PNS (DF)	PDAD
Sexual orientation	3,35%	4,90%	0,64%
Gender identity	-	-	0,17%

Note: "PNS (DF)" represents the equivalent rate obtained from the National Health Survey (PNS), considering only data from the Federal District (DF)

Further insights come from the logistic regression models presented in Table 2, which estimate the odds ratios (ORs) of being in the nonresponse group for SOGI questions, controlling for key sociodemographic factors. Results from the PNS 2019 show that women were significantly more likely than men to refuse or not know how to answer the sexual orientation question (OR = 1.3, $p < 0.01$). Also, younger individuals (18–29) were the reference group, and older age groups, especially those 65+, were substantially less likely to be nonrespondents (OR = 0.43, $p < 0.001$). People with higher education were associated with a reduced likelihood of nonresponse: those with a college degree had a 47% lower chance of refusing (OR = 0.53, $p < 0.01$). Notably, individuals who refused to answer other sensitive items, such as age at first sexual intercourse (OR = 13.15, $p < 0.001$) or sexual activity in the past 12 months (OR = 1.92, $p < 0.001$), were significantly more likely to also refuse the SOGI question—suggesting that item nonresponse may reflect a broader discomfort with personal topics.

In the PDAD 2021, predictors of nonresponse followed similar patterns but with key distinctions like married, cohabiting, divorced, and widowed individuals were significantly less likely to be in the nonresponse group for both sexual orientation and gender identity questions—suggesting that social stability or life-course stage may enhance disclosure. People assigned female at birth were less likely to refuse the gender identity question (OR = 0.48, $p < 0.01$), reinforcing hypotheses that men or male-identified proxies may be more hesitant to disclose gender-diverse identities.

For gender identity, older age groups (especially 40–49 and 50–64) were significantly more likely to be nonrespondents compared to the 18–29 group. Interestingly, individuals with elementary education had lower odds of refusing to answer the gender identity question compared to those with no education (OR = 0.24, $p < 0.10$), although education was not a significant predictor for sexual orientation nonresponse in PDAD. Finally, refusal to disclose monthly income was associated with higher odds of nonresponse to both SOGI items, suggesting a broader pattern of selective disclosure in proxy-based interviews.

These preliminary results reinforce the need to consider survey context, data collection methodology, and individual characteristics when designing and interpreting SOGI measures. The consistent association between nonresponse to SOGI and other sensitive items underscores the role of privacy, perceived risk, and stigma in shaping respondents' willingness to disclose. Moreover, while proxy responses may improve item completion rates, they may do so at the cost of authentic representation, particularly for LGBTQ+ individuals whose identities are not always visible to other household members.

Table 2: Odds Ratio for Nonresponse to Sexual Orientation and Gender Identity Questions, PNS 2019, PDAD 2021

Variables	PNS 2019	PDAD 2021	
	SO	SO	GI
Race (White-Yellow Ref.)			
Black-Brown-Indigenous	1.20	1.37*	1,38
Sex (Masculine Ref.)			
Feminine	1.3**	1.11	0,48**
Age group (18-29 years Ref.)			
30-39 years	0.59***	1.43*	1,10
40-49 years	0.65**	1.40	3,27***
50-64 years	0.75	1.46	2,25*
65 +	0.43***	0.19*	0,86
Education (< elementary Ref.)			
Elementary	1.13	1.55	0,24*
High school	0.81	1.54	0,68
College Degree	0.53**	1.02	0,80
Marital Status (Single Ref.)			
Cohabitation	1.15	0.47*	0,23*
Married	0.91	0.19***	0,24***
Divorced	0.59**	0.45*	0,26*
Widowed	0.83	0.11*	0,00***
Employment status (Employed Ref.)			
Unemployed	1.22	1.09	1,80
Answer regarding the age at first sexual relation (Answered Ref.)			

Don't know/Don't remember	6.9***	-	-
	13.15***	-	-
Answer regarding sexual relations in the last 12 months (Answered Ref.)			
Refused to answer	1.92***	-	-
Answered to earnings from any source (Answered Ref.)			
Refused to answer	-	1.55***	1,60

Note: *** p-value <0,001; ** p-value <0,01; * p-value <0,10; Ref. = reference group

Discussion and Future Steps

The results presented in this study highlight how nonresponse to SOGI questions is not merely a technical artifact, but rather a socially and methodologically mediated outcome. The sharp contrast in nonresponse rates between the two Brazilian surveys—PNS (2019) and PDAD (2021)—demonstrates the substantial impact that survey design and data collection method have on identity disclosure. While proxy reporting appears to reduce item nonresponse dramatically, it also raises concerns regarding data validity, as proxies may not accurately report the sexual orientation or gender identity of other household members, either due to lack of knowledge or reluctance to disclose sensitive information.

These findings align with international research suggesting that contextual, cognitive, and emotional factors influence disclosure behavior (Rosser et al., 2021; Suen et al., 2020). Respondents may refuse to answer SOGI questions because of perceived privacy threats, anticipated stigma, or confusion about terminology. In the Brazilian case, where gender and sexuality continue to be politicized topics, and where household-based data collection is the norm, these dynamics are especially pronounced. The elevated nonresponse rate in the self-reported PNS likely reflects these barriers more directly, while the low refusal rate in the proxy-based PDAD may obscure deeper silences and erasures within the data.

Importantly, our results suggest that nonresponse is not randomly distributed across the population. In both surveys, certain demographic groups—such as older adults, individuals with lower education, men, and those who refuse to answer other sensitive questions—are more likely to be in the nonresponse group. These patterns are consistent with studies from high-income countries (e.g., Cho, 2023; Kim & Fredriksen-Goldsen, 2013) and reinforce the need to interpret nonresponse as a socially structured phenomenon, embedded in life-course trajectories, social position, and perceived risk of disclosure.

Another key finding is the strong association between refusal to answer SOGI items and refusal to respond to other intimate questions, such as those related to sexual behavior or income. This supports the hypothesis that nonresponse may reflect a broader disposition toward withholding personal information in survey contexts—a form of informational self-protection (Fredriksen-Goldsen & Kim, 2015). In this sense, item nonresponse itself becomes meaningful: it can indicate the limits of what respondents are willing (or able) to share, depending on who is asking, how, and under what conditions.

These insights point to several future directions for research, data collection, and policy:

- Survey design: National statistical agencies should consider testing different formats for SOGI questions (e.g., open-ended vs. closed, one-step vs. two-step), and evaluating the trade-offs between self and proxy reporting. Greater investment in privacy-enhancing techniques and interviewer training may reduce nonresponse without sacrificing data validity.
- Community engagement: Collaborating with LGBTQ+ organizations and advocacy groups in the formulation of questions and in the dissemination of results can improve trust and cultural sensitivity, especially among historically marginalized subgroups such as trans, non-binary, and black people.
- Policy relevance: Strengthening the collection of SOGI data is essential not only for demographic accuracy but also for evidence-based policymaking. As Brazil expands the inclusion of SOGI questions in national surveys, monitoring nonresponse patterns will be key to ensuring the representativeness and reliability of findings used to guide health, education, and labor policies.
- Comparative research: Further cross-national studies—particularly across Latin America—could help uncover how legal contexts, cultural norms, and survey institutions shape patterns of SOGI disclosure and erasure. Comparative designs would allow researchers to assess whether the Brazilian experience is unique or part of broader regional dynamics.

In sum, this study contributes to the growing field of queer demography by showing that nonresponse is not just a limitation to be corrected, but a phenomenon that can reveal the boundaries of legibility and recognition in survey systems shaped by cisnormative and heteronormative logics. Addressing these challenges requires not only technical adjustments but also a political commitment to visibility, equity, and inclusion in official statistics.

References

Carpenter, C. S., Eppink, S. T., Gonzales, G., & McKay, T. (2021). Effects of access to legal same-sex marriage on marriage and health. Journal of Policy Analysis and Management, 40(2), 376-411.

Cho, S. (2023). Socio-demographic Determinants of Item Nonresponse to a Question on Sexual Orientation: Evidence from the National Health Interview Survey. International Journal of Contents, 19(4).

Drydakis, N. (2022). Sexual orientation and earnings: a meta-analysis 2012–2020. Journal of Population Economics, 35(2), 409-440.

Fredriksen-Goldsen, K. I., & Kim, H. J. (2015). Count me in: Response to sexual orientation measures among older adults. Research on Aging, 37(5), 464-480.

Suen, L. W., Lunn, M. R., Katuzny, K., et al. (2020). What sexual and gender minority people want researchers to know about sexual orientation and gender identity questions: A qualitative study. Archives of Sexual Behavior, 49(7), 2301–2318.

Kim, H. J., & Fredriksen-Goldsen, K. I. (2013). Nonresponse to a question on self-identified sexual orientation in a public health survey and its relationship to race and ethnicity. American journal of public health, 103(1), 67-69.

Lee, S., Fredriksen-Goldsen, K. I., McClain, C., Kim, H. J., & Suzer-Gurtekin, Z. T. (2018). Are sexual minorities less likely to participate in surveys? An examination of proxy nonresponse measures and associated biases with sexual orientation in a population-based health survey. Field Methods, 30(3), 208-224.

LeVasseur, M. T., Kelvin, E. A., & Grosskopf, N. A. (2013). Intersecting identities and the association between bullying and suicide attempt among New York city youths: results from the 2009 New York city youth risk behavior survey. American journal of public health, 103(6), 1082-1089.

Pinto, I. V., Andrade, S. S. D. A., Rodrigues, L. L., Santos, M. A. S., Marinho, M. M. A., Benício, L. A., ... & Canavese, D. (2020). Perfil das notificações de violências em lésbicas, gays, bissexuais, travestis e transexuais registradas no Sistema de Informação de Agravos de Notificação, Brasil, 2015 a 2017. Revista Brasileira de Epidemiologia, 23(suppl 1), e200006-SUPL.

Rosser, B. R. S., Levesque, J., Greene, G. J., Iantaffi, A., & Smolenski, D. J. (2021). Acceptability and feasibility of collecting sexual orientation and expanded gender identity data in urology and oncology clinics. LGBT Health, 8(6), 420–426.

Suen, L. W., Lunn, M. R., Katuzny, K., et al. (2020). What sexual and gender minority people want researchers to know about sexual orientation and gender identity questions: A qualitative study. Archives of Sexual Behavior, 49(7), 2301–2318.

Suliano, D., Jesus Filho, J. D., & Irffi, G. (2021). Sexual orientation and wage differentials using anthropometric and health measures. Estudos Econômicos (São Paulo), 51, 111-142.

Young, L. J., & Rater, B. R. (2021). The Farm Producer Survey: Unit and Item Nonresponse (No. 1496-2022-1759).

Zhang, Z., Smith-Johnson, M., & Tumin, D. (2023). Contextual Influences on Nonresponse to Health Survey Questions About Sexual Orientation and Gender Identity. LGBT health.