TECHNICAL NOTE Nº IDB-TN-03376

Uncovering the Hidden: Innovative Methods for Data Collection from Hard-to-Reach Populations

Carina Lupica
Yasmin Mertehikian
Sofía Caravedo
Christian Silvera

Inter-American Development Bank
Gender and Diversity

August 2026

Uncovering the Hidden: Innovative Methods for Data Collection from Hard-to-Reach Populations

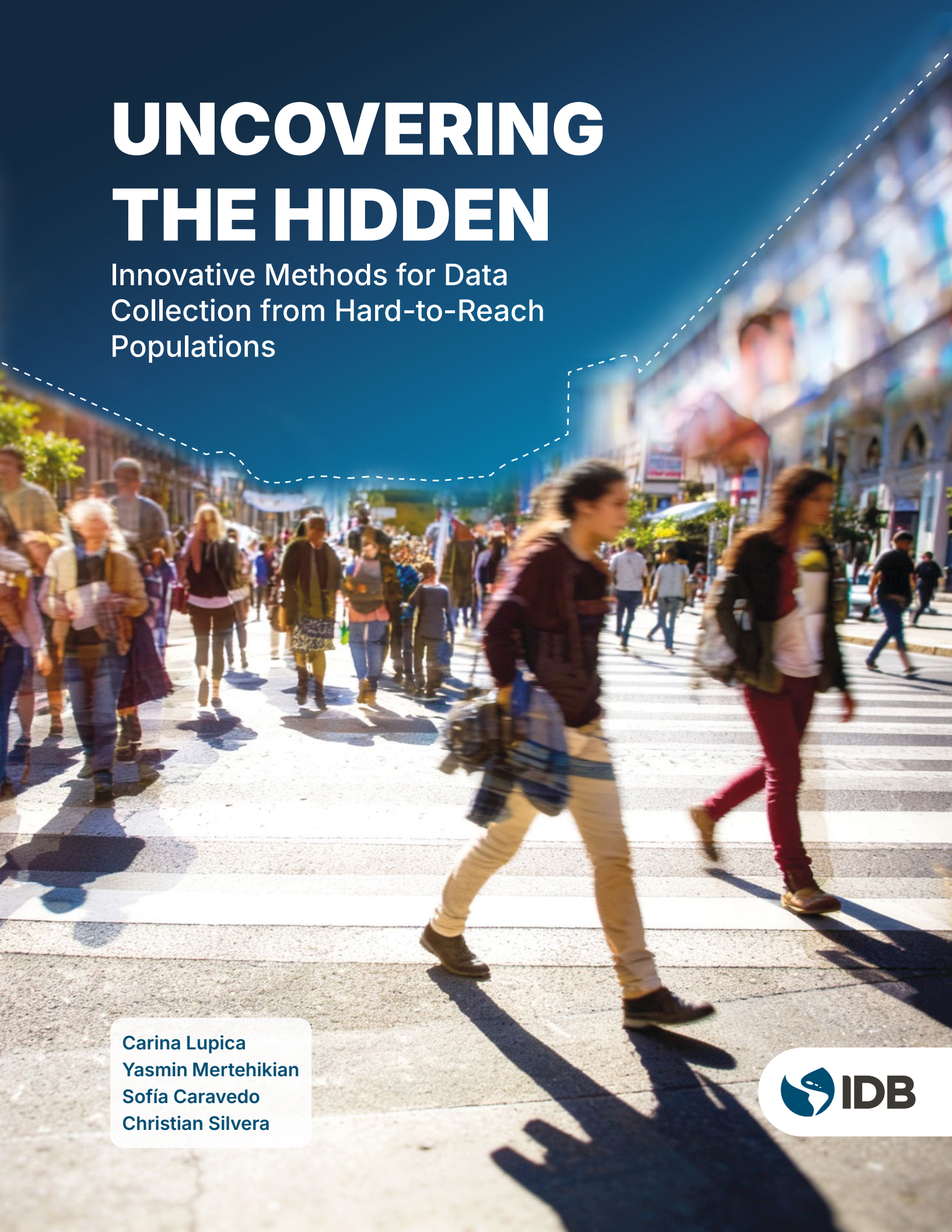
Carina Lupica
Yasmin Mertehikian
Sofía Caravedo
Christian Silvera

Inter-American Development Bank
Gender and Diversity

August 2026

UNCOVERING THE HIDDEN

Innovative Methods for Data Collection from Hard-to-Reach Populations



Carina Lupica
Yasmin Mertehikian
Sofía Caravedo
Christian Silvera

IDB

Inter American Development Bank
1300 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org

Content

UNCOVERING THE HIDDEN

Innovative Methods for Data Collection for Hard-to-Reach Populations

Carina Lupica, Yasmin Mertehikian, Sofía Caravedo y Christian Silvera[1]

Abstract

This document examines the challenges associated with data collection from hard-to-reach populations, which are characterized by limited statistical visibility due to structural, social, legal, or stigma-related barriers. These limitations create information gaps that directly affect the quality of public policy design, implementation, and evaluation, as well as the efficient allocation of resources and the ability to promote inclusive growth.

Drawing on a review of the literature and applied experiences, this document systematizes and compares a range of innovative methods used to study these populations, many of which originated in public health and were subsequently adapted for social research. Among the most commonly used approaches are Time-Location Sampling (TLS), Targeted Sampling (TS), Respondent-Driven Sampling (RDS), and Snowball Sampling, each with its own strengths and limitations depending on the context, the density of social networks, and the research objectives. The document also highlights the value of complementing these methods with qualitative techniques, such as Mystery Shopping and Vignette Studies, to capture deeper dimensions of social experience.

The document emphasizes that no single method is applicable to all situations; therefore, methodological selection should be guided by the context and the characteristics of each population. In practice, the involvement of community members and the development of trust-based relationships with key stakeholders are essential for improving access and data quality, alongside ethical considerations such as anonymization, the use of respectful language, and the creation of safe environments. Ultimately, research involving these populations requires balancing the generation of evidence with the ethical responsibility not to deepen existing rights violations, recognizing that the most effective approaches combine methodological pluralism, contextual sensitivity, and a commitment to inclusion and to making historically marginalized realities visible.

JEL Classification: C80, C81, C82, C83, J18

Key words: *hard-to-reach populations, research design, sampling methods, public policy*

1. Carina Lupica and Yasmin Mertehikian work in the Gender and Diversity Division at the IDB, and Sofía Caravedo and Christian Silvera work at Equilibrium Business, Data and Communities. The authors sincerely thank Suzanne Duryea, Luciana Etcheverry, Nadin Medellín, Marta Luzes, Ercio Muñoz Saavedra, Blanca Torrinco, Yanira Marcela Oviedo-Gil, Luz Orellana Bautista, María Gutiérrez Damas, Yren Rotela Ramírez, Juan Alonzo González, Lisbeth Morales Pernalete, Juanita Ardila Hidalgo, and Lisa G. Johnston for their valuable collaboration and for sharing their expertise during interviews about their research experiences using different methods for data collection from hard-to-reach populations. The authors also extend their gratitude to Cecilia Martinez Gomez, Francisco Riera, Rocío Ccasani, Milagros Reyes, and Jackeline Velarde for their assistance and support in editing this document. Andrea Iturra, from Words for Development, designed the publication. The cover image was created with the support of Magnific's artificial intelligence tools. Any errors or omissions are the sole responsibility of the authors. The contents and conclusions of this document reflect the views of the authors and do not necessarily represent those of the IDB, its Board of Executive Directors, or the countries they represent.

Introduction

Hidden populations pose significant challenges for research and public policy design. These challenges stem primarily from the absence of sampling frames that would allow these populations to be identified and from the limited access to their members. The literature provides various definitions that, collectively, share two central characteristics: first, the inability to construct lists or sampling frames that would enable their selection through traditional sampling methods; and second, the existence of barriers that hinder access to their members. These barriers may be associated with social factors—such as health or employment status—, cultural or legal factors, or processes of stigmatization (Heckathorn, 1997; Shaghaghi et al., 2011).

Hidden populations include, for example, people who use illegal drugs, sex workers, people with irregular migration status, people experiencing homelessness, people living with HIV, incarcerated individuals, LGBTQ+ people, and survivors of domestic or sexual violence, among others. Although these groups are diverse, they share characteristics that limit their statistical visibility and their participation in conventional research studies.

Many of the methods[2] used to study these populations were initially developed in the field of public health. In that context, understanding disease transmission, risk behaviors, and inequalities in access to health services was essential. Over time, however, these methods have also proven to be valuable tools for analyzing the social, cultural, and economic dynamics that shape the experiences and living conditions of these populations.

In this context, there is a need to compile and systematize the various methods used for data collection from hard-to-reach populations, as well as to analyze the lessons learned from their design and implementation. Reviewing previous experiences—both in academic research and in the field of public policy—makes it possible to identify the most appropriate approaches for addressing the characteristics, needs, and contexts of these populations.

To address the question of **how to select the most appropriate method** based on research or policy objectives, this document provides an overview of the different approaches used to study hidden populations. It presents their main strengths and limitations, illustrates them with practical examples, and offers practical recommendations based on the literature and the accumulated experience of researchers and practitioners working with these populations.

2. The use of terms such as method and technique in this document should not be confused with their meaning in the fields of epistemology, philosophy of science, and the study of scientific methods. In these disciplines, the concepts of method, methodology, technique, and research instrument are not equivalent, and using them interchangeably may create confusion. In this document, method refers to the logical steps that structure the generation of scientific knowledge (from the formulation of the research problem and research question to the development and testing of hypotheses) and may be classified, for example, as deductive or inductive, depending on the direction of reasoning in relation to the data. Methodology, in contrast, is understood here as the manner in which that method is operationalized, guiding the type of data and analyses that drive the research process. At this level, distinctions are made among quantitative, qualitative, and mixed-methods approaches, depending on whether the emphasis is placed on statistical generalization from large samples or on understanding processes and mechanisms through smaller samples. Techniques refer to the specific procedures used to collect and analyze information (e.g., surveys, interviews, observation, or experiments), whereas instruments are the specific tools used for those purposes, including questionnaires, measurement scales, structured interview guides, standardized tests, among others. The issue of population sampling falls within the level of instrument design, as it encompasses the strategies through which individuals are identified and selected to provide the information needed to answer the research questions. However, for ease of reading, the terms method, technique, and sampling instrument are used interchangeably throughout this document.

1 Theoretical and Conceptual Framework

Theoretical Perspectives and Definitions

There are several definitions of hidden populations (see TABLE 1). In general terms, the concept refers to **groups with low social visibility that are difficult to access and for which no lists, registries, or sampling frames exist to enable their identification and selection through conventional sampling methods**. This situation hinders not only their identification and enumeration but also the production of reliable information on their characteristics and living conditions.

The study of hidden populations requires recognizing a set of distinguishing characteristics that largely explain the methodological and ethical challenges associated with their research. In these contexts, technical challenges—such as access, participant recruitment, and the collection of sensitive information—cannot be viewed in isolation from ethical considerations, as they directly affect the protection of participants' autonomy, safety, and confidentiality[3].

Table 1. Hidden Populations: Definitions and Characteristics

Source/Year	Definition
Watters & Biernacki, 1989	Hidden populations are groups that are socially invisible because some of their practices are clandestine, placing them beyond public view and outside the reach of formal mechanisms of social control.
Heckathorn, 1997	These groups share two defining characteristics: **(1)** they lack a sampling frame, meaning that the size of the population is unknown; and **(2)** they are concerned about being identified because their activities or identities are socially stigmatized, leading many individuals to decline participation in research studies or to provide less reliable responses in order to protect their privacy.
UNAIDS & WHO, 2010	The World Health Organization (WHO) defines these as groups of individuals who, because some of their HIV-related risk behaviors are illegal or socially stigmatized, tend to remain hidden. This situation makes them difficult to reach through conventional epidemiological surveillance methods.
Shaghaghi, Bhopal & Sheikh, 2011	Hidden populations are characterized by being inaccessible through conventional survey methods because of their low social visibility and the absence of available sampling frames. These groups often exhibit some of the following characteristics: **(1)** high mobility, **(2)** a tendency to conceal their activities for fear of confrontation with authorities, and **(3)** social pressure exerted by majority groups.
Vidal, 2021	Although the International Organization for Migration (IOM) does not have its own definition of a hidden population, it recognizes that many migrants may belong to these groups, making them difficult to identify and enumerate. It also notes that migrants in the most vulnerable situations rarely appear in official statistics, contributing to the limited availability of high-quality migration data.

Source: Prepared by the authors.

3. Research involving human participants entails an inherent tension between protecting individual well-being and generating knowledge for the public good. See Resnik, D. B. (2018). *The Ethics of Research with Human Subjects: Protecting People, Advancing Science, Promoting Trust.* Springer International Publishing.

Characteristics of Hidden Populations

Hidden or hard-to-reach populations often exhibit characteristics such as high mobility, concealment strategies driven by social stigma or fear of legal sanctions, and experiences of discrimination and social rejection that discourage them from publicly identifying themselves. These conditions not only result in a scarcity of statistical information but also limit the development of evidence-based public policy, reducing the capacity of institutions to respond effectively to their needs. The specialized literature has identified at least six key characteristics that help explain the complexity of researching these populations.

Lack of a Sampling Frame (C1)

Hidden populations are not covered by official lists or registries[4] that would allow them to be identified systematically (Heckathorn, 1997). The absence of a sampling frame prevents the use of conventional sampling methods—such as simple random sampling, stratified sampling, or cluster sampling—which require a preexisting reference frame to calculate representative samples (Cochran, 1977). To address this limitation, researchers rely on alternative methods, such as Snowball Sampling or Respondent-Driven Sampling (RDS)[5], which provide valid approaches, albeit with certain limitations regarding representativeness and statistical validity (Heckathorn, 1997).

Tendency to Conceal One's Status or Activities (C2)

Members of these populations often avoid disclosing their status or activities to researchers, authorities, or other institutions because of fear of legal or social sanctions. This tendency results in a lower willingness to participate in surveys and interviews, as well as the provision of incomplete or inaccurate information (Shaghaghi et al., 2011). This dynamic operates at both the individual and collective levels, as members of these groups seek to protect not only their own safety but also that of other group members.

Stigmatization and Social Exclusion (C3)

The characteristics, behaviors, or circumstances associated with these populations are often stigmatized, resulting in their social exclusion (Watters and Biernacki, 1989). This situation may create fear of judgment, rejection, or criminalization, constituting an additional barrier to accessing essential services and support programs. Social exclusion and stigmatization can increase their vulnerability and compromise their overall well-being (UNAIDS and WHO, 2010).

4. This refers to censuses, population registers, or other institutional records that make it possible to identify the members of a population.
5. Method known as Respondent-Driven Sampling (RDS).

Situation of Vulnerability (C4)

Hidden populations often live under conditions of vulnerability, understood as the lack of the resources and opportunities necessary to fully exercise their rights and meet their basic needs. This includes both tangible deprivations—such as limited access to health care, education, housing, or employment—and forms of social exclusion and the absence of institutional recognition (WHO and UN, 2010).

High Mobility and Geographic Dispersion (C5)

In many cases, hidden populations do not remain in a single location but instead move frequently, making it difficult to monitor and follow them over time (Heckathorn, 1997). This mobility may be internal—within a city or region—or external—across countries—making it challenging to establish stable points of contact for research (Vidal, 2021). In addition, geographic dispersion increases the costs and logistical resources required for sampling and data collection.

Low Trust in Institutions (C6)

Hidden populations often exhibit distrust toward government institutions, research centers, and even some civil society organizations. This attitude is associated with previous experiences of discrimination, harassment, or persecution, leading many individuals to hesitate or feel reluctant to provide information or participate in research studies (Shaghaghi et al., 2011). For this reason, building trust and strengthening relationships with the community are essential through practices that safeguard personal information, minimize risks, and ensure transparency in the collection and use of data. These principles are fundamental to the ethics of research involving human participants and are necessary to promote legitimacy and participation in contexts of vulnerability (Resnik, 2018).

Groups Identified as Hidden Populations

Various authors have documented the presence of hidden populations across different contexts, noting that these groups often remain invisible because of legal, social, or health-related conditions that give rise to stigma or exclusion. **TABLE 2** summarizes this literature by identifying the groups considered to be hidden populations, the authors who have studied them, and the characteristics they exhibit according to the typology described in the previous sections (C1–C6).

Table 2. Comparative Classification of Hidden Populations

Hidden Population	Author(s)	C1	C2	C3	C4	C5	C6
People who use illegal drugs	Watters and Biernacki (1989)	✔	✔	✔	✔		✔
Sex workers	Heckathorn (1997)	✔	✔	✔	✔		✔
Irregular migrants	Shaghaghi, Bhopal and Sheikh (2011)		✔		✔	✔	✔
People experiencing homelessness	Shaghaghi, Bhopal and Sheikh (2011)	✔		✔	✔		✔
People living with HIV/AIDS	UNAIDS and WHO (2010)		✔	✔	✔		✔
Men who have sex with men (MSM)	UNAIDS and WHO (2010)		✔	✔	✔		✔
LGBTQ+ people	Equilibrium Business, Data and Communities (BDC) (2025)	✔	✔	✔	✔		✔
People working in illegal economies, such as smuggling	Heckathorn (1997)	✔	✔	✔		✔	✔
People with certain diseases (e.g., COVID-19)	Soltanian et al. (2020)	✔	✔	✔	✔		

Source: Prepared by the authors based on Watters and Biernacki (1989); Heckathorn (1997); Shaghaghi, Bhopal, and Sheikh (2011); UNAIDS and WHO (2010); Equilibrium (2023); and Soltanian et al. (2020).

TABLE 2 shows that most hidden populations exhibit multiple characteristics that make them difficult to identify and include in research. The following examples illustrate these dynamics:

- **People who use illegal drugs and sex workers** exhibit nearly all of the characteristics analyzed (C1, C2, C3, C4, and C6), reflecting high levels of social invisibility and pronounced distrust of institutions.
- **Irregular migrants** are characterized primarily by their high mobility (C5), which makes them difficult to locate and creates significant gaps in administrative records and statistical data.
- **People experiencing homelessness** share the absence of official records (C1) and conditions of structural vulnerability (C4), limiting their representation in conventional research.
- **People living with HIV/AIDS and men who have sex with men (MSM)** primarily face barriers associated with social stigmatization (C3) and distrust of institutions (C6).
- The COVID-19 pandemic demonstrated that communicable diseases can give rise to hidden populations by combining the concealment of health status (C2) with the absence of sampling frames (C1) and fear of stigmatization (C3).

Taken together, this analysis shows that these populations are not isolated within broader patterns of exclusion but instead share a set of characteristics that reinforce their invisibility and deepen their social vulnerability.

Finally, it is important to note that contextual factors—such as geographic location, the social and cultural environment, the historical moment, and the prevailing legal framework and public policy—significantly influence how these populations are perceived, recognized, and studied. Accordingly, the proposed classification should not be understood as a rigid or universal typology, but rather as an analytical framework grounded in the literature and open to variation according to the specific conditions of each context.

2 Methods for Researching Hidden Populations

Research involving hidden populations presents a methodological challenge because of the absence of formal sampling frames and the existence of structural barriers that hinder access to their members. In particular, the application of traditional probability sampling methods is constrained by at least two assumptions that do not hold in these contexts.

First, these methods assume the existence of complete and accessible sampling frames—such as administrative records or official databases—that make it possible to identify and randomly select units of analysis. However, in the case of hidden populations, such sampling frames are often nonexistent or incomplete (Heckathorn, 1997; Shaghaghi et al., 2011).

Second, they assume that individuals are willing to participate openly and to be identified as members of the study population. However, many hidden populations face barriers to interacting with formal systems of registration or identification because their characteristics, practices, or circumstances are often associated with contexts of stigmatization, criminalization, or social sanction.

To overcome these limitations, the methodological literature has developed a range of innovative sampling techniques, both probability-based and non-probability-based, that enable researchers to better reach these populations. These methods seek to balance two key objectives: **(i)** expanding access to hard-to-reach groups and **(ii)** improving the validity and representativeness of the data collected.

The following sections present seven methods commonly used in research involving hidden populations. The methods are organized according to their methodological and operational complexity (from least to most complex), that is, the level of technical structure, statistical requirements, and logistical procedures needed to access these populations or estimate their characteristics.

The first five methods correspond to different strategies designed to identify, access, and estimate the characteristics of populations with low social visibility or for which clearly defined sampling frames do not exist.

In addition, two complementary methods used for data collection among hidden populations (Methods 6 and 7) are included. These approaches focus primarily on observing and exploring experiences, practices, and perceptions, providing descriptive and contextual information that is particularly useful for understanding barriers, interaction dynamics, and behavioral patterns.

Snowball Sampling

Description

Snowball Sampling is a non-probability sampling method in which the initial participants in a study—referred to as the initial sample or "seeds"—refer other individuals who belong to the target population, creating a chain-referral recruitment process. The method leverages existing social connections or networks within the population of interest to reach individuals who may be difficult to contact through conventional sampling methods, such as simple random sampling or stratified sampling. The literature indicates that Snowball Sampling is useful in both qualitative and quantitative research, particularly in contexts characterized by low social visibility or high levels of stigmatization (Ting et al., 2025).

Sample Selection

This method requires a clear definition of the target population and the establishment of rigorous inclusion criteria. Data collection begins with a seed sample consisting of participants intentionally selected by the research team. Each participant refers other individuals who meet the predefined inclusion criteria, who in turn may recruit additional participants, resulting in a chain-referral recruitment process.

Recruitment continues until the target sample size is reached or theoretical saturation is achieved, defined as the point at which the inclusion of additional participants no longer contributes new properties or theoretical insights relevant to the analysis (Bryant and Charmaz, 2007) [6]. Careful monitoring and regulation of the referral process are essential, as the method is particularly susceptible to the formation of clusters that may produce an overly homogeneous sample or one composed of highly similar individuals. In this regard, Saunders et al. (2023) note that regulating referrals contributes to a more diverse and balanced sample, thereby strengthening the validity and accuracy of the data collected.

In certain contexts, incentives may be provided both for direct participation and for referring or recruiting additional participants into the sample. Finally, the diversity and representativeness of the resulting sample are assessed by analyzing the demographic and social characteristics of the surveyed population (Ting et al., 2025).

6. Hennink and Kaiser (2022) conclude that sample saturation depends primarily on the richness of the information collected rather than on the total number of participants.

Figure 1. Snowball Sampling **Process**

Source: Ting et al., 2025. Prepared by the authors.

Strengths of the Method

- Provides broad access to hidden populations through social networks and trusted relationships.
- Is efficient in terms of time and cost compared with probability sampling methods.
- Promotes high response rates by relying on trusted relationships among participants.

Methodological Limitations

- Offers limited control over the size and composition of the sample, which may introduce homophily bias—that is, the tendency of individuals to connect with others who share similar characteristics.
- May result in representativeness issues, as certain subgroups of the population tend to be overrepresented while others are underrepresented or excluded.
- Requires larger sample sizes than probability-based methods to mitigate its inherent limitations.

Snowball Sampling is a valuable strategy for studying hidden populations because it facilitates access to hard-to-reach groups through trusted relationships and social networks. This technique has also been shown to generate high response rates due to the privacy and support provided by participants' social networks. Nevertheless, its limitations with respect to representativeness and the potential for bias indicate that its findings should be interpreted with caution. In this regard, the technique is particularly useful as an exploratory tool or as a complement to other methods, rather than as a stand-alone approach for producing generalizable estimates of the target population.

CASE STUDY 01 - Nova Southeastern University: Doctoral Students' Use of LinkedIn to Access Hidden Populations

Dusek et al. (2015) document the use of online professional networks, particularly LinkedIn, as a tool for recruiting participants in studies involving hard-to-reach populations. In this case study conducted at Nova Southeastern University, doctoral students combined Targeted Sampling (TS) and snowball sampling by distributing online questionnaires through personalized links shared on the platform.

The use of personalized links made it possible to initiate recruitment through existing professional contacts and progressively expand the sample through referral networks. This approach facilitated access to participants who would otherwise have been difficult to identify or contact and achieved relatively high response rates (42–54%). In addition, the online format offered advantages such as greater privacy when responding to sensitive questions and a significant reduction in the costs associated with data collection.

However, the authors also note several methodological limitations. In particular, reliance on existing professional networks may bias recruitment toward certain participant profiles; chain-referral recruitment may limit sample diversity; and the circulation of survey links across broad networks reduces control over questionnaire dissemination, creating challenges for sample representativeness and data quality control.

Source: Dusek et al. (2015).

Targeted Sampling

Description

Targeted Sampling (TS) is a non-probability sampling method based on the strategic selection of geographic areas and specific subgroups within the target population. Its application relies on the research team's prior knowledge and the completion of an initial mapping or ethnographic assessment. The method involves identifying and documenting specific populations within defined geographic areas and subsequently developing a systematic recruitment plan to ensure the inclusion of an adequate number of participants from each group or location (Watters and Biernacki, 1989).

Sample Selection

The process begins with the identification of relevant subgroups within the target population based on an ethnographic assessment or preliminary diagnostic that characterizes their main sociodemographic characteristics, patterns of interaction, and geographic distribution.

These subgroups are then transformed into sampling strata, with each stratum assigned a specific quota according to its relative size or representativeness within the overall population. Within each stratum, a predetermined number of participants who meet the study's inclusion criteria are selected. This procedure contributes to more balanced and systematic coverage of the different segments of the population under study (Watters and Biernacki, 1989).

Strengths of the Method

- Expands coverage of the most difficult-to-reach subgroups within hidden populations.
- Enables researchers to focus efforts and optimize resources by prioritizing strategically selected locations.
- Allows for adjustments during implementation, such as redirecting recruitment toward underrepresented areas or subgroups.

Methodological Limitations

- Depends heavily on the completeness and accuracy of the initial assessment or diagnostic.
- May increase costs and operational demands (human resources and time), particularly when subgroups are geographically isolated or dispersed.
- Is susceptible to biases associated with site selection, which may limit the generalizability of findings if not all areas where the target population is concentrated are included.

Compared with Snowball Sampling, Targeted Sampling (TS) represents an advancement by incorporating more systematic planning and broader coverage of subgroups within hidden populations. Because it is based on an initial ethnographic assessment, it enables more efficient allocation of resources and helps improve the representativeness of the different segments under study. However, its effectiveness depends

on the quality of the preliminary assessment and on the operational capacity to cover diverse geographic areas and subgroups, resulting in greater demands in terms of time and cost.

Consequently, this method is a useful alternative for studies seeking to overcome the exploratory limitations of chain-referral sampling while still remaining significantly dependent on the quality of the initial mapping of key subgroups and locations.

CASE STUDY 02 - Recruiting Participants in Urban Settings Through Geographic Mapping in Baltimore, Maryland

Allen et al. (2019) applied Targeted Sampling (TS) to collect data from street-based sex workers and examine the relationship between policing practices, health outcomes, and HIV exposure. Using geographic mapping, the research team identified strategic areas where the target population could be located.

In the first stage, secondary data—including prostitution arrest records, 911 calls, and online forum posts—were compiled to determine the spatial distribution of sex work. Based on this information, the researchers developed a preliminary map of neighborhoods, selecting those with the highest concentration of incidents and the greatest diversity of meeting locations (downtown, peripheral areas, and mixed residential-commercial zones). Field visits were then conducted through vehicle patrols, walking surveys, and ride-alongs with police officers to validate the identified active locations.

In the second stage, a sampling frame was constructed based on the selected areas. Researchers conducted 129 field shifts and more than 500 hours of fieldwork, resulting in the recruitment of 250 sex workers. During implementation, several sampling sites had to be adjusted because some locations proved too large or exhibited low levels of activity during the recruitment periods.

Targeted Sampling (TS) proved to be a flexible and effective approach for accessing hidden populations in urban settings, provided that it is accompanied by rigorous spatial analysis to appropriately define sampling units (a combination of location, day, and time) and capture temporal variability. Its principal challenges include high resource requirements, limitations in representativeness, and the potential exclusion of subgroups located in peripheral or hard-to-reach areas.

Source: Allen et al., 2019.

Capture-recapture

Description

The capture-recapture method originated in ecology, where it was developed to estimate the size of animal populations in their natural habitats. It was subsequently adapted for use in the social and health sciences to estimate the size of hidden or hard-to-reach populations.

It is a non-probability method based on conducting two or more rounds of "capture" involving individuals from the target population. During the first round, participants are identified and recorded, while in subsequent rounds, new individuals are recorded and those who have already been captured are identified. The validity of the estimates depends on the fulfillment of several critical assumptions, most notably the stability of the population's size and composition throughout the observation period (Shaghaghi et al., 2011).

Sample Selection

1. The process begins with an initial capture phase, during which individuals belonging to the target population are identified and recorded (C_1).
2. This is followed by a second recapture phase, in which individuals from the same population are identified again, with particular attention given to those who were already captured during the first phase (C_2).
3. Based on the information obtained from both rounds, the Lincoln–Petersen estimator is applied to estimate the total population size (N_e) using the relationship between the number of individuals captured in the first round, those captured in the second round, and those recaptured in both rounds (Adebajo et al., 2013).

Shaghaghi et al. (2011) present the Lincoln–Petersen formula as follows:

$$N_e = \frac{(C_1 + 1)(C_2 + 1)}{(R + 1)} - 1 \quad (1)$$

Where:

- N_e Estimated population size.
- C_1 Number of individuals captured during the first capture.
- C_2 Number of individuals captured during the second capture.
- R Number of individuals who were recaptured (i.e., appeared in both samples).

Applying this formula makes it possible to estimate the total size of the hidden population under study. Although this estimate is approximate and depends on the assumptions underlying the method, it provides valuable information for assessing the size of the target population, guiding the design of research strategies, and planning interventions that contribute to the development of public policy better aligned with the realities of that population.

Strengths of the Method

- Enables estimation of the size of the target population without requiring a sampling frame.
- Is particularly useful for studying groups concentrated in specific locations.
- Is relatively simple and cost-effective to implement compared with other population estimation methods.

Methodological Limitations

- Relies on strict assumptions, particularly population stability between capture rounds (i.e., no migration, mortality, or addition of new members), a condition that is rarely fully met in social settings.
- Is less effective for geographically dispersed or highly mobile populations.
- May introduce recapture bias, as individuals do not necessarily have the same probability of being identified or captured again.

The capture-recapture method is a valuable tool for generating quantitative estimates of the size of hidden populations without requiring a sampling frame. Its principal strengths include its relative operational simplicity, low cost, and feasibility in settings where the target population is geographically concentrated. However, the accuracy of its estimates depends heavily on the fulfillment of strict assumptions, particularly population stability and equal probabilities of recapture, which limits its applicability in more dynamic social contexts.

CASE STUDY 03 - Application of the Capture-Recapture Method to Estimate the Size of a Population at High Risk of HIV in Nigeria

Adebajo et al. (2013) used the capture-recapture method to estimate the size of the population of men who have sex with men engaged in sex work (MSM-SW) in three Nigerian cities: Lagos, Kano, and Port Harcourt. Researchers identified 56 meeting locations in Kano, 38 in Lagos, and 42 in Port Harcourt where MSM-SW met with clients. The findings indicated that, on an average weekend night, the estimated population was approximately 723 in Port Harcourt, 620 in Lagos, and 353 in Kano.

The study highlighted the existence of a population at high risk of HIV that remains largely invisible due to the stigmatization and criminalization of both homosexuality and sex work. It also underscored the need for prevention programs and research specifically targeting these populations.

Implementation Steps:

1. **Key informant interviews:** Conducted to identify the locations and times at which MSM-SW met with their clients.
2. **First capture:** MSM-SW participants were interviewed and each was given a "marker" (in this case, a keychain).
3. **Second capture:** Two weeks later, researchers returned to the same locations at the same times, repeated the interviews, and asked participants whether they had received a keychain during the first capture.
4. **Recording matches:** Researchers recorded the number of participants captured during the first and second rounds, as well as the number who had received keychains (i.e., those captured and recaptured).
5. **Population size estimation:** The Lincoln–Petersen formula was applied to estimate the population size in each city and calculate the corresponding confidence intervals.

Source: Adebajo et al. (2013).

Time-Location Sampling

Description

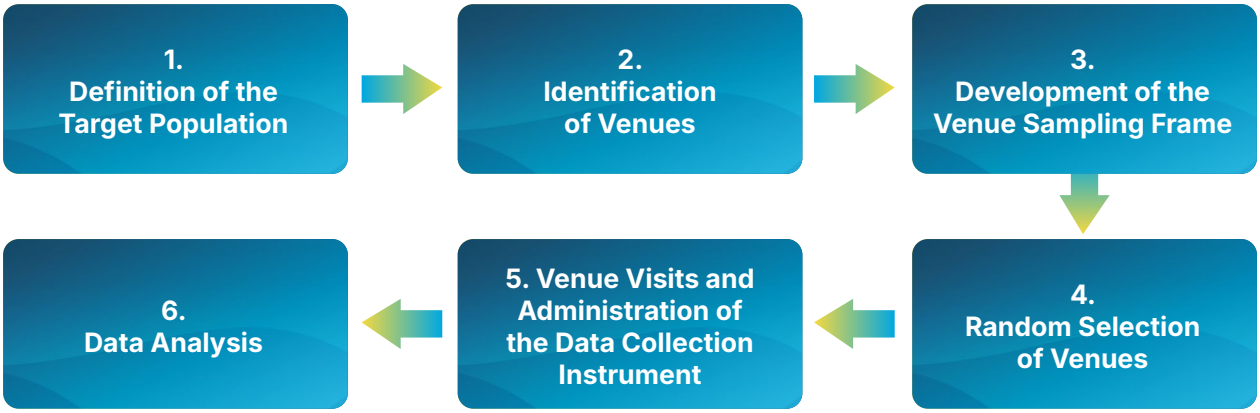
Time-Location Sampling (TLS) is a probability sampling method used to collect data from hidden populations by sampling the locations where the target population regularly gathers. In this approach, the initial sampling units are venues, defined as specific combinations of location, day, and time period during which members of the target population regularly congregate.

The random selection of venues makes it possible to construct an operational sampling frame and organize the systematic recruitment of participants. However, population coverage may be limited when certain members of the target population do not attend the identified venues (Karon and Wejnert, 2012).

Sample Selection

1. **Exploratory phase:** A mapping of the venues frequented by the target population is conducted. This information is used to construct the sampling frame by identifying the relevant locations and time periods.
2. **Probability-based selection of venues:** Venues are selected using simple random sampling or stratified sampling, with venues serving as the primary sampling units. This approach ensures representative coverage of the locations where the target population is concentrated.
3. **Systematic recruitment:** Visits to the selected venues are scheduled, and data collection instruments are administered systematically (for example, by inviting every third eligible individual entering the venue during the specified time period to participate).
4. **Data analysis:** The collected data are analyzed using sampling weights based on the probability of selection in order to generate estimates that are representative of the population attending the venues.

Figure 2. Time-Location Sampling **Process**



Source: Prepared by the authors, based on Karon and Wejnert, 2012.

Strengths of the Method

- Provides a rigorous theoretical foundation that is compatible with traditional probability survey methods.
- Reduces masking bias and the limitations associated with chain-referral sampling.
- Facilitates the application of advanced statistical models, such as linear and logistic regression.

Methodological Limitations

- Excludes individuals who do not frequent the selected locations or venues, resulting in non-coverage bias.
- Generally involves higher operational costs than other methods because of the duration, planning, and complexity of fieldwork.
- Has limited representativeness, as inferences are restricted to the population attending the venues included in the sampling frame.

Time-Location Sampling (TLS) provides greater statistical rigor by relying on probability sampling principles and a structured sampling frame based on venues. Its application helps reduce masking bias and facilitates more systematic and comparable analyses. However, its coverage is limited to individuals who attend the identified venues, excluding those who do not frequent these locations. Although TLS represents an advancement in terms of validity and comparability, its implementation requires careful definition of the sampling frame and the availability of sufficient resources to support extensive fieldwork.

CASE STUDY 04 - Application of Time-Location Sampling (TLS) for Behavioral Surveillance in Guatemala

The Division of Global HIV/AIDS, Center for Global Health, at the Centers for Disease Control and Prevention (CDC) in Atlanta, in collaboration with the Universidad del Valle de Guatemala, implemented Time-Location Sampling (TLS) in behavioral surveillance studies involving men who have sex with men (MSM).

An initial inventory of venues frequented by the target population was developed using information from previous studies, key informant interviews, and focus groups. Researchers mapped all public and semi-public locations where MSM socialized or sought partners, including nightclubs, bars, gyms, and other venues.

The validity and accessibility of each location were subsequently verified through exploratory site visits, which were used to construct an updated sampling frame of active venues. Each month, approximately 40 venues were selected at random, while an additional two venues were purposively selected by the study staff to ensure representation of areas or types of venues that were visited less frequently. Venues with fewer than seven eligible participants were replaced by others with similar characteristics.

At each venue, participants were approached systematically (every fifth individual), written informed consent was obtained, the survey questionnaire was administered, and educational materials together with a USD$6 incentive were provided. Researchers also recorded both participant enrollment and refusals to participate.

Time-Location Sampling (TLS) offers advantages such as a strong theoretical foundation and the ability to conduct statistical analyses using widely available software packages (e.g., Python, RStudio, or Stata). However, the study emphasizes the importance of maintaining consistency in sample representation, as any variation may affect the composition of the sample and the ability to assess trends over time.

Source: Paz-Bailey et al., 2013.

Respondent-Driven Sampling

Description

Respondent-Driven Sampling (RDS) is a participant recruitment method based on chains of referrals among members of the target population. Although it begins as a non-probability sampling method, its combination with mathematical weighting and adjustment models allows for the correction of biases resulting from nonrandom selection. As a result, RDS acquires characteristics similar to those of probability sampling, increasing its validity for empirical research and facilitating the estimation of population parameters.

RDS is organized through controlled recruitment chains using specific mechanisms that regulate and facilitate the enrollment of new participants. Its main components are as follows:

- **Dual incentive system:** Each participant receives one incentive for participating and another for recruiting peers.[7]
- **Limitation on the number of referrals:** Each participant may invite only a predetermined number of individuals.
- **Chain-referral recruitment:** New participants are recruited through referrals from existing participants rather than directly by the research team (Heckathorn, 1997, 2002).

Sample Selection

The process begins with a group of participants known as "seeds," who are purposively selected because of their access to and connections within the target population.[8] Each seed receives a limited number of coupons[9] to recruit new participants. These recruits, in turn, receive the same number of coupons to continue the recruitment process through successive waves,[10] until the desired sample size is reached.

7. In some cases, this mechanism is also used in studies employing snowball sampling.
8. Seeds are members of the target population who should possess three key characteristics: they should be highly motivated, diverse from one another, and have extensive social networks (Raifman et al., 2022).
9. A coupon is the instrument provided to each participant to invite a limited number of peers to join the study. Although the number of coupons distributed to each participant is not fixed, most studies use between two and three coupons in practice to balance the pace of recruitment and prevent the overrepresentation of particular subgroups.
10. Waves are the successive rounds of recruitment in which participants recruit new participants.

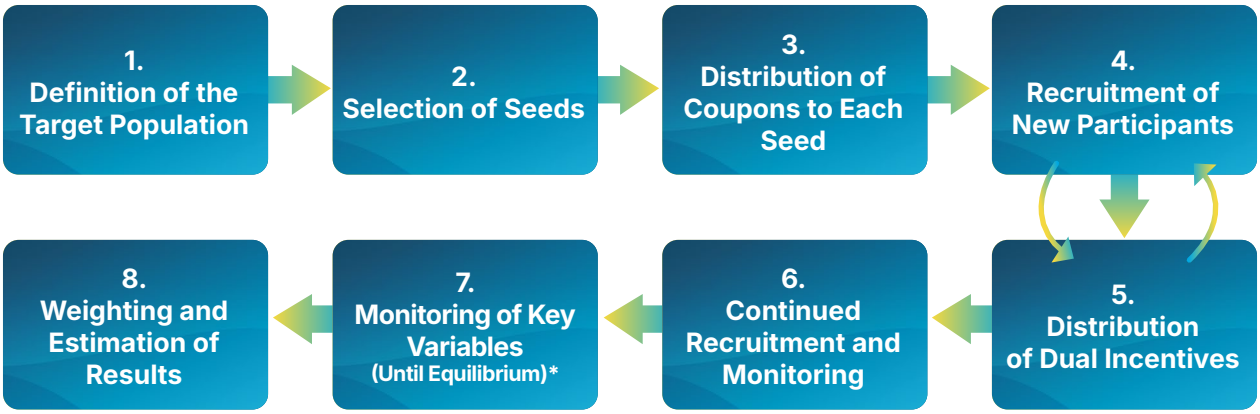
This mechanism allows researchers to:

- Gradually achieve statistical equilibrium, observed when the proportions of key variables (e.g., sex, age, and geographic location) no longer change substantially across successive recruitment waves.
- Increase sample diversity, as reflected in the progressive inclusion of subgroups of interest within the target population, such as different age groups, sexual orientations, or socioeconomic levels.

At the same time, information is collected on the size of each participant's personal network.[11] These data are used to weight the sample and correct for biases resulting from nonrandom selection. Specifically, because individuals with larger social networks are more likely to be recruited, inverse weights are applied based on network size: participants with larger networks receive lower weights, whereas those with smaller networks receive higher weights.

The process continues until the planned sample size is reached or statistical equilibrium in the sample composition is achieved—that is, when the main demographic characteristics stabilize across successive recruitment waves (Heckathorn, 1997, 2004).

Figure 3. Respondent-Driven Sampling **Process**



*Refers to the point at which the sample results no longer change substantially as additional participants are recruited; that is, when the main characteristics (e.g., age, gender, or educational attainment) remain stable.

Source: Prepared by the authors, based on Heckathorn, 1997, 2002.

11. Each participant is asked to report how many individuals within the target population they know, according to the criteria established by the research team.

Strengths of the Method

- Enables the production of population-representative estimates through weighting factors derived from the size of participants' personal networks.
- Reduces certain selection biases associated with chain-referral recruitment, including the overrepresentation or underrepresentation of specific groups.
- Provides an efficient and dynamic recruitment process, as the dual-incentive system encourages participation and helps increase both the size and diversity of the sample.

Methodological Limitations

- Depends on the functioning of the target population's social networks and therefore requires participants with large, active, and diverse personal networks.
- Inadequate selection of seeds may affect the initial composition of the sample and delay or prevent the achievement of statistical equilibrium.
- Involves substantial analytical complexity: the calculation of sampling weights, population estimators, and design effects[12] requires specialized technical expertise and appropriate statistical software.

Respondent-Driven Sampling combines the logic of chain-referral recruitment with statistical methods designed to approximate the properties of probability sampling. Its ability to generate more representative estimates makes it a robust alternative to other non-probability sampling methods, provided that seeds are appropriately selected and the research team has the technical expertise required for data processing. Nevertheless, its reliance on active social networks and its considerable analytical complexity mean that successful implementation depends not only on methodological rigor but also on the operational capacity of the research team.

12. The design effect (DEFF) measures the extent to which the variance of an estimator increases as a result of using a complex sampling design, compared with a Simple Random Sample (SRS) of the same size. In methods such as Respondent-Driven Sampling (RDS), factors including dependence among observations, homophily within social networks, and the use of sampling weights increase the DEFF, thereby reducing the precision of the estimates. This represents a methodological limitation because it requires larger sample sizes and specialized statistical procedures to calculate valid standard errors and confidence intervals.

CASE STUDY 05 - Characterizing Undocumented Venezuelan Migrants Using Respondent-Driven Sampling

The World Bank, in collaboration with the Colombia office of the NGO Innovations for Poverty Action (IPA), implemented a project targeting Venezuelan migrants in Ecuador and Peru to characterize their vulnerabilities and living conditions. To achieve this objective, the study employed Respondent-Driven Sampling to obtain a sample that approximated representativeness among undocumented migrants, a population characterized by high levels of distrust toward institutions, particularly in the context of the COVID-19 pandemic.

The study consisted of three components.

1. **Exploratory interviews:** Researchers conducted 42 semi-structured interviews to identify trust-related concerns, such as fear of deportation.
2. **Consultation with local organizations:** Meetings were held with community organizations working on migration-related issues.
3. **RDS data collection:** The study began with a small group of seeds, each of whom was asked to provide a predetermined number of referrals. This process continued through successive recruitment waves until a total of 3,455 migrant households had been recruited.

The sampling frame combined information from migrant organizations with referrals provided by study participants, allowing the sample to be updated dynamically over time. The study also highlighted the value of the dual-incentive system (one incentive for participation and another for referrals) as an effective strategy for encouraging participation among hard-to-reach populations.

Challenges Identified:

- **Distrust of the recruitment process:** Stemming from previous experiences with scams and fraudulent messages distributed through WhatsApp, Facebook, and telephone calls.
- **Adaptation of the survey instrument:** Modifying language and terminology to reflect local usage (e.g., using municipality instead of city and homologation instead of validation), as well as adjusting the form of address (from the formal *usted* to the informal *tú*) to improve clarity and naturalness.
- **Support for field staff:** IPA trained interviewers to prevent vicarious trauma—the emotional impact of repeatedly hearing distressing experiences—through psychological support sessions and training in empathy and emotional coping strategies.

Source: Rozo, S. (2021).

Mystery Shopping

Description

Mystery shopping is a covert evaluation method in which shoppers—members of the target population who have been trained in advance—pose as actual customers. Its purpose is to collect direct, objective, and standardized information on service quality, compliance with established protocols, interactions between service providers and users, and other key aspects of the customer experience.

This method is characterized by its ability to recreate real-world service encounters, allowing researchers to compare observed experiences against quality standards established in advance by the research team. Unlike other evaluation tools, Mystery Shopping provides real-time data from the user's perspective, making it a valuable instrument for quality management, staff training, and continuous service improvement (Allison, Severt, and Dickson, 2010).

In studies involving hidden populations, this technique has been used to document how public and private institutions interact with these groups. Examples include an evaluation of the quality of HIV/STI testing services at 43 clinics in southeastern Michigan (Bauermeister et al., 2015) and a study in Mexico that applied the Mystery Shopping method in 761 establishments—including pharmacies and clinics—to assess the quality of sexual and reproductive health services provided to adolescents (De Castro et al., 2018).

Selection of Shoppers

The process of selecting and organizing shoppers consists of several stages designed to ensure the objectivity, standardization, and relevance of the observations:

1. **Identification of establishments:** A list of establishments of interest (e.g., clinics, pharmacies, public institutions, or commercial businesses) is compiled. The status of each establishment is verified through official records, online searches, or telephone calls to confirm the services offered, costs, and operating hours.
2. **Selection and training of shoppers:** Shoppers are selected based on characteristics that closely match those of the target population (e.g., age, gender, or other relevant attributes). Training includes standardized scripts, role-playing exercises, and practice interactions.
3. **Assignment of visits:** Shoppers are assigned to establishments according to a design intended to capture variation in service delivery (e.g., different days and times). During each visit, the shopper follows a predefined scenario and records observations objectively in accordance with the methodological criteria established for the study.
4. **Documentation and validation:** After each interaction, shoppers complete a standardized assessment form evaluating dimensions such as accessibility, privacy, staff interactions, and other qualitative observations. The research team subsequently conducts debriefing sessions to consolidate and validate the information collected.

Strengths of the Method

- Enables the observation of service delivery under real-world conditions.
- Produces standardized experiences that facilitate comparisons across different services or establishments.
- Generates both quantitative and qualitative information on the quality of service delivery.

Methodological Limitations

- Has limited external validity, as the findings cannot be generalized to the entire user population.
- May be affected by biases related to shopper performance if shoppers do not adhere strictly to the standardized scripts and protocols.
- Raises ethical considerations related to identity simulation and the covert collection of information or evaluation of services.

Mystery Shopping is a valuable tool for evaluating the quality of services provided to hidden populations under real-world conditions. Its ability to generate practical evidence on gaps in service delivery makes it a useful resource for improving programs and public policy. However, its limitations in terms of representativeness and the ethical implications associated with identity simulation suggest that it should be used as a complement to other methods rather than as a primary research strategy.

CASE STUDY 06 - Evaluating the Quality of HIV/STI Services Through Mystery Shopping

Bauermeister et al. (2015) applied the Mystery Shopping method in 43 HIV/STI clinics across southeastern Michigan to evaluate the cultural competence and quality of services provided to young gay and bisexual men. Five shoppers were trained, and a standardized assessment instrument covering 13 dimensions was used: accessibility, wait time, confidentiality, privacy, cleanliness, availability of informational materials, comfort of the facility, staff interactions, use of inclusive language, respect for identity, quality of counseling, clarity of information, and perceived safety.

Implementation Steps:

1. **Study design and training:** Shoppers with characteristics similar to those of the target population were selected and trained, including simulations of the HIV/STI testing experience.
2. **Field implementation:** Each clinic was visited by two mystery shoppers, who requested routine HIV/STI testing services without disclosing their role as evaluators.
3. **Documentation of experiences:** After each visit, shoppers completed a structured questionnaire covering the 13 evaluation dimensions.
4. **Comparative analysis:** Scores were analyzed by comparing clinics that offered HIV testing only (n = 14) with those providing comprehensive HIV/STI services (n = 29).
5. **Results and interpretation:** Overall scores were positive; however, gaps were identified in counseling and sexual health education.

The study suggests that Mystery Shopping is an effective method for identifying opportunities to improve sensitive aspects of service delivery, such as staff interactions, confidentiality, and counseling. However, its usefulness may be influenced by the inherent subjectivity involved in evaluating personal service experiences.

Source: Bauermeister et al. (2015).

Vignette Studies

Description

Vignette Studies are a qualitative method that involves presenting participants with hypothetical scenarios or visual, auditory, or combined representations designed to elicit responses to specific situations. Their purpose is to explore perceptions, judgments, attitudes, or potential behaviors in response to particular contexts (McInroy and Beer, 2022).

The process begins with defining the target population and developing hypothetical scenarios that represent situations relevant to the phenomenon under study. These vignettes may be presented in various formats, including text, images, videos, or other digital media. Before implementation, it is recommended that the vignettes undergo a validation process—such as expert review or pilot testing—to ensure their clarity, cultural relevance, and appropriate interpretation by participants (McInroy and Beer, 2022).

During implementation, participants respond to structured questions or engage in guided discussions based on the scenarios presented. The data collected may be analyzed using qualitative approaches—exploring meanings, perceptions, or narratives—or quantitative approaches, by comparing judgments, attitudes, or decisions across different groups of participants.

Strengths of the Method

- Allows researchers to explore attitudes and reactions to sensitive topics without directly exposing participants, as it relies on hypothetical scenarios.
- Ensures that all participants respond to equivalent stimuli, facilitating the comparability of responses.
- Offers high methodological flexibility, as it can be implemented in multiple formats (e.g., text, images, videos, or online platforms), extending its applicability across diverse populations and settings.

Methodological Limitations

- Responses obtained may not necessarily reflect participants' actual behavior in real-world situations.
- Even minor variations in the wording, tone, or format of the vignettes may introduce response bias.
- Findings are often highly dependent on the context and the characteristics of the population under study, limiting their generalizability to broader populations.

CASE STUDY 07 - Ethical Considerations in the Use of Deception in Research

Methods that employ simulated agents, such as Mystery Shopping and Vignette Studies, may require the use of deception as a methodological strategy to obtain information on perceptions or behaviors in contexts where revealing the true purpose of the study could influence participants' responses. Deception may take explicit forms—through the provision of false or manipulated information—or implicit forms, through the concealment or omission of relevant information.

The regulation of deception varies across disciplines. In sociology and psychology, strict ethical codes establish criteria governing its use, including those developed by the American Psychological Association, the American Sociological Association, and ESOMAR's guidelines for Mystery Shopping. In economics, the American Economic Association (2018) also provides guidance on the use of deception. Moreover, experimental economics has developed a strong social norm discouraging the use of deception, primarily because of its implications for data validity and participant trust.

Before implementing methods that involve deception, researchers should carefully evaluate the following considerations:

- **Importance:** The use of deception should be justified by the scientific, educational, or applied value of the study.
- **No harm:** The strategy should not cause physical, social, or psychological harm to participants, including significant emotional distress.
- **Alternatives:** Researchers should demonstrate that the objectives of the study cannot be achieved using alternative methods. For example, deception may be warranted when it is necessary to control variation in the stimulus or to prevent participants from modifying their behavior in order to present themselves in a socially desirable manner.
- **Costs and resources:** Researchers should assess the implementation costs relative to alternative methodological approaches that may be impractical because of their complexity or logistical requirements.

Best Practices

- **Debriefing:** Whenever possible, participants should be informed of the study's true objectives after their participation and given the opportunity to withdraw their data.
- **Voluntary participation and anonymity:** These are fundamental principles that should be respected throughout the research process.
- **Ethical review:** The use of deception should be reviewed and approved by an Institutional Review Board (IRB).
- **Systematic documentation:** Because subjective judgments may be involved, researchers should define clear, observable categories—whether structural or behavioral—to ensure consistency and transparency in data collection.

Source: Espinosa & Brañas—Garza, 2024 y Morena de Diago, 2013.

Comparative Analysis of Methods

Following the review of the various methods used to study hidden populations, the table below presents a comparison of their principal characteristics and areas of application. To facilitate this comparison, five analytical criteria were defined to systematically assess the strengths and limitations of each method:

- **Methodological rigor:** The degree of the method's theoretical, mathematical, and statistical foundation, as well as the consistency and appropriateness of its underlying assumptions in relation to the specific objectives of the technique.
- **Coverage:** The method's ability to reach both geographically concentrated and dispersed populations.
- **Analytical depth:** The level of detail and quality of the information that can be obtained about the characteristics of the study population.
- **Flexibility and applicability:** The extent to which the method can be adapted to different social contexts, population types, and available resources.
- **Representativeness:** The method's ability to produce estimates that accurately reflect the target population and support the generalization of findings, when such generalization is consistent with the objectives and nature of the technique.

For this comparison, a standardized rating scale ranging from 1 to 5 was applied, defined as follows:

- **1 (Very Low):** The method meets the criterion to a very limited extent and faces significant challenges in its application.
- **2 (Low):** The method partially meets the criterion but has important limitations that affect its performance.
- **3 (Moderate):** The method adequately meets the criterion but requires certain adjustments or improvements to strengthen its consistency and effectiveness.
- **4 (High):** The method meets the criterion well, demonstrating strong performance across most application contexts.
- **5 (Very High):** The method fully satisfies the criterion, achieving high standards of robustness and reliability.

In some cases, decimal values (e.g., 2.5 or 3.5) were assigned to reflect intermediate assessments between two points on the scale. It should be noted that the assigned scores represent an interpretive assessment derived from a process of critical reflection, informed by a review of the specialized literature and a comparative analysis of the strengths and limitations reported in previous studies.

Table 3. Assessment of Methods for Researching Hidden Populations

Method	Methodological Rigor	Coverage	Analytical Depth	Flexibility/ Applicability	Validity and Reliability
Snowball Sampling	3,5	3,0	2,5	3,5	3,0
Targeted Sampling	2,5	2,5	2,5	3,5	3,0
Capture-Recapture	2,5	2,0	2,0	2,5	3,0
Time-Location Sampling, TLS	4,0	3,5	3,0	3,0	4,0
Respondent-Driven Sampling, RDS	4,0	4,0	3,0	3,0	4,0
Mystery Shopping	2,0	1,5	4,0	4,0	3,0
Vignette Studies	2,0	1,0	3,5	4,0	2,5

Source: Prepared by the authors.

Key Findings from the Comparative Analysis

- **Methodological rigor:** Respondent-Driven Sampling (RDS) and Time-Location Sampling (TLS) are the most well-established methods, supported by stronger theoretical, mathematical, and statistical foundations. This results in more structured study designs and better-defined assumptions for producing population estimates. By contrast, although Snowball Sampling, Targeted Sampling (TS), and the capture-recapture method are valuable for exploratory research, they have a lower degree of theoretical and statistical formalization, which limits the robustness of their procedures and assumptions compared with more structured approaches.
- **Coverage:** RDS stands out for its ability to reach broad and relatively dispersed social networks. Targeted Sampling (TS), in contrast, is better suited to populations concentrated in specific locations or subgroups. Methods such as capture-recapture, Mystery Shopping, and Vignette Studies generally provide more limited coverage, reducing their usefulness in settings characterized by high geographic or social heterogeneity.
- **Analytical depth:** Qualitative and mixed-method approaches, such as Mystery Shopping and Vignette Studies, provide richer and more detailed information on participants' experiences, perceptions, and behaviors, often exceeding statistically more robust methods in terms of descriptive depth. In contrast, methods such as RDS and TLS, while producing more reliable population estimates, typically provide less in-depth information on social processes or individual-level dynamics.
- **Flexibility and applicability:** The suitability of each method depends on several factors, including the study objectives, implementation context, available budget, characteristics of the hidden population (such as size, mobility, or degree of concealment), ethical considerations, and the availability of prior information. Nevertheless, Targeted Sampling (TS) and Mystery Shopping tend to be more adaptable to different contexts and operational constraints, making them particularly useful for field research with logistical limitations. In contrast, the capture-recapture method requires stricter assumptions—such as the relative stability of the population—which may limit its applicability in certain settings.
- **Validity and reliability:** RDS and TLS have the strongest empirical support, producing results that are generally more consistent, replicable, and representative.

In summary, social network-based methods, such as Respondent-Driven Sampling (RDS), and venue-based methods, such as Time-Location Sampling (TLS), offer significant advantages when the objective is to produce more robust and potentially generalizable population estimates. In contrast, exploratory and qualitative approaches—such as Snowball Sampling, Mystery Shopping, and Vignette Studies—are particularly valuable because of their greater methodological flexibility, fewer requirements for strict statistical assumptions, and ability to provide deeper insights into the social dynamics, perceptions, and experiences of hidden populations.

Ethical Considerations

Research involving hidden populations should incorporate the fundamental ethical principles applicable to research with human participants, including do no harm, confidentiality, data anonymization, and the use of respectful and appropriate language. It should also adhere to international ethical protocols established by the relevant authorities and be reviewed by Institutional Review Boards (IRBs) that are independent of the research team. Informed consent should be understood both in its formal dimension—as a documented agreement—and in its substantive function of ensuring that participants fully understand the objectives, scope, and potential implications of their participation.

Issues requiring special attention when working with hidden populations:

1. **Strengthening the capacity of field teams.** It is essential to provide training and awareness-building workshops for field staff. A recommended practice is to include members of the study population on field teams, use active learning techniques such as role-playing to assess participants' understanding and application of training content, and promote ongoing feedback and support throughout the data collection process.
2. **Strengthening confidentiality protocols.** Surveys, interviews, and other research activities should be conducted in safe, private settings to minimize the risk of exposing participants. This precaution is particularly important when access to the population is limited because of safety concerns or stigma. Recommended practices include innovative techniques that reduce direct interaction with the interviewer, such as using audio-recorded questions delivered through headphones, allowing participants to respond independently and confidentially.
3. **Validating instruments and processes with members of the study population.** Involving members of the target population in the design and validation of research instruments improves both data quality and the legitimacy of the research process. This approach helps identify which questions are most relevant for obtaining meaningful information, ensures that they are phrased in a context-sensitive manner, and incorporates practical and cultural considerations that facilitate their implementation.

Table 4. Summary of Methods for Researching Hidden Populations

Method Name	Sampling Type	Description	Sample Selection	Advantages	Limitations
Snowball Sampling	Non-Probabillistic	Chain-referral recruitment in which initial participants refer other members of the target population.	Purposive selection of an initial group; each participant refers additional individuals until the desired sample size is reached.	Broad reach; cost- and time-efficient; high response rates due to trust among referred participants.	Limited control over sample composition; homophily bias (the tendency for individuals to connect with others who share similar characteristics); representativeness concerns.
Targeted Sampling	Non-Probabillistic	Strategic selection of geographic areas and subgroups based on a prior assessment.	Identification of relevant subgroups; participation quotas are assigned according to each subgroup's estimated proportion of the population.	Facilitates the identification and coverage of hidden subgroups; allows resources to be focused on strategic locations; can be adjusted during implementation.	Depends on the quality of the initial assessment; entails higher operational costs; risk of bias if not all areas where the target population is located are included.
Capture-Recapture	Non-Probabillistic	Estimates population size by comparing individuals identified repeatedly across two or more rounds of data collection.	Initial capture of individuals during the first round, followed by recapture during a second round; estimates are based on the comparison between rounds.	Allows estimation of population size; simple and cost-effective; useful for geographically concentrated populations.	Requires population stability; less effective for dispersed populations; susceptible to bias if adequate recapture is not achieved.
Time-Location Sampling, TLS	Probabillistic	Random selection of combinations of locations, days, and times where the target population gathers.	Identification of venues (specific combinations of location, day, and time); random selection of venues; systematic recruitment at each selected venue.	Provides statistical rigor; reduces concealment bias by covering multiple location-time combinations; supports advanced statistical analyses.	Excludes individuals who do not attend venues; more costly and operationally complex; representativeness is limited to the defined sampling frame.

Respondent-Driven Sampling, RDS	Non-probabilistic with quasi-probabilistic statistical adjustments	An enhanced version of Snowball Sampling that incorporates incentives and statistical weighting.	Identification of seeds with access to the target population; wave-based recruitment using coupons; application of statistical weights to adjust for selection bias.	Can produce representative population estimates; can correct for selection bias; facilitates efficient participant recruitment.	Depends on active and well-connected social networks; risk of inadequate seed selection; complex analysis requiring advanced technical expertise.
Mystery Shopping	Mixed (qualitative and quantitative)	Covert evaluation conducted by trained shoppers who pose as real users or customers to assess service quality, protocols, and staff interactions.	Identification of establishments; selection and training of shoppers with characteristics similar to those of the study population; assignment of visits across different days and times; systematic recording of observations.	Enables observation of services under real-world conditions; produces standardized experiences that facilitate comparison; generates both qualitative and quantitative data.	Findings are not generalizable; risk of bias if shoppers do not follow the established protocols; subject to ethical considerations related to identity simulation and covert evaluation.
Vignette Studies	Mixed (qualitative and quantitative)	Present participants with hypothetical scenarios (e.g., text, images, videos) to explore perceptions, attitudes, or judgments regarding sensitive or specific situations.	Development and prior validation of the vignettes; recruitment of participants, who respond to or discuss the proposed scenarios.	Allow sensitive topics to be explored without directly revealing the study objective; ensure comparability because all participants respond to the same stimuli; highly versatile (text, images, video, online formats).	Responses may not reflect actual behavior; small changes in vignette design may introduce bias; findings are context-dependent and difficult to generalize.

3 Lessons from Practice: What Should Be Considered When Researching Hidden Populations?

The purpose of this section is to provide practical guidance for researching hard-to-reach populations, based on findings and lessons learned from qualitative interviews with researchers who have used the methods described in the previous section. As part of this study, 16 individuals with experience in methodological design, decision-making, implementation, and fieldwork were interviewed. Participants included representatives from academia, civil society organizations, consulting firms, community-based organizations, multilateral institutions, as well as members of hidden populations, including LGBTQ+ individuals and people in situations of human mobility.[13]

All testimonies were anonymized and focused on methodological decisions, fieldwork challenges, and lessons learned. The semi-structured interviews were conducted virtually in accordance with protocols on confidentiality, anonymity, and informed consent, including compliance with the Personal Data Privacy Policy and the Data Subject Request Form, in accordance with Inter-American Development Bank (IDB) requirements.

Based on this information, the following section presents practical recommendations for strengthening ethically sound and methodologically rigorous research involving hidden populations.

13. Human mobility refers to the movement of people within a country or across international borders, regardless of whether the movement is voluntary or forced, temporary or permanent. The concept encompasses various forms of mobility, including labor migration, internal displacement, refugee movements, and other forms of movement in contexts of vulnerability (International Organization for Migration. (2020). *Glossary on Migration* (IML No. 34). IOM.).

1. Approaching Hidden Populations

The interviews suggest that, when researching hidden populations, it is useful to distinguish between **hard-to-reach populations and hard-to-measure populations**.

> "I would distinguish between populations that are hard to reach and those that are hard to measure, because some of these populations are difficult to locate, while in others the issue is that we interview them all the time, but the topics are sensitive and people are unwilling to disclose their identity or these behaviors."
>
> Anonymized testimony. Interview series conducted for this publication.

Hard-to-reach populations face barriers such as geographic dispersion, limited institutional contact, or membership in closed social networks. In contrast, hard-to-measure populations are characterized by challenges related to the sensitive nature of the information being collected or individuals' reluctance to disclose it. This distinction helps guide methodological design and supports the development of tailored strategies based on the specific factors contributing to the hidden nature of the population.

2. Research Design

The interviews indicate that research involving hidden populations requires a gradual, flexible, and context-specific design that combines methodological rigor with sensitivity toward historically stigmatized or hard-to-reach groups. The experiences shared by the interviewees highlight three complementary approaches that enable researchers to adapt methods to contexts characterized by distrust, low visibility, and the absence of sampling frames.

INITIAL ASSESSMENT OF THE POPULATION

Before selecting research methods, it is essential to identify what is known—and what remains unknown—about the target population. This includes reviewing existing administrative records, census data, and documentary sources, as well as assessing factors such as social connectivity, geographic dispersion, topic sensitivity, and levels of distrust toward institutions. This initial assessment helps anticipate fieldwork risks, identify information gaps, and design mixed-method approaches that combine qualitative and quantitative techniques tailored to the specific characteristics of each population.

> **"[...] We always begin with an initial assessment. What do we know, and what do we not know about the population? What are its channels of organization? What are the risks involved in approaching the population? That is where we start designing mixed-method approaches, but over time we have learned that these methods need to be adapted."**
>
> Anonymized testimony. Interview series conducted for this publication.

BUILDING A FOUNDATIONAL UNDERSTANDING OF THE HIDDEN POPULATION

The interviews suggest that, once the initial assessment of the target population has been completed, the next step is to prepare the research team and verify the underlying methodological assumptions. This process should focus on two main areas:

- Preliminary formative research: It is recommended to examine the structure of the group's social networks by identifying strong and weak ties, subgroups, and potential bottlenecks that could affect representativeness in sampling methods such as Respondent-Driven Sampling (RDS). This helps refine the methodological approach before implementation.
- Research team training: All team members should share a common language, understand the research protocols, and respect diverse identities, thereby strengthening both the quality and validity of the data collected.

> "We review all of these processes so that certain parameters are clearly understood even within the research team itself. That is ultimately reflected in the research, but it is important that the very first step be this training, so that we all speak the same language."
>
> Anonymized testimony. Interview series conducted for this publication.

IDENTIFYING KEY STAKEHOLDERS

The interviews indicate that identifying community leaders and institutional partners is essential for building trust and facilitating access to the field. This process is typically gradual and progresses in stages, beginning with in-depth interviews with key stakeholders, followed by focus groups or group interviews. Institutional partnerships provide operational and financial support, help align the research with concrete needs, and enhance the relevance of the findings.

> "When the population is this hidden, we have to work in phases. It almost always begins with identifying a key person within the population, conducting an in-depth interview, and then moving on to focus groups or group interviews."
>
> Anonymized testimony. Interview series conducted for this publication.

Taken together, these three approaches demonstrate that fieldwork decisions begin during the research design phase and that technical planning must be integrated with a thorough understanding of the research context. According to the experiences shared by the interviewees, this approach makes it possible to conduct research that is feasible, ethical, and responsive to the characteristics of the populations involved, while improving the quality, relevance, and reliability of the data collected.

3. Field Team

The interviews highlight that access to hidden populations depends largely on the composition, preparation, and cultural sensitivity of the team responsible for data collection. Beyond the research instruments or logistical arrangements, the quality of interactions with participants is critical for building trust and obtaining reliable information. Based on the experiences shared by the interviewees, the following recommendations emerged:

INCLUDE MEMBERS OF THE TARGET POPULATION ON THE RESEARCH TEAM

Including members of the target population—as interviewers, facilitators, or methodological advisors—facilitates access to the field, improves understanding of cultural norms, and strengthens trust throughout the data collection process. This practice not only promotes equity but also serves as a methodological strategy that enhances the quality of the information collected.

> **"A cisgender person should not be interviewing a transgender person about their life experiences, because there are many things that will simply be lost. [...] That initial foundation of trust is extremely important."**
>
> Anonymized testimony. Interview series conducted for this publication.

PARTNER WITH COMMUNITY ORGANIZATIONS AND LOCAL STAKEHOLDERS

Partnerships with civil society organizations and community leaders help identify key contacts, review research instruments, and adapt fieldwork strategies. These organizations also play a critical role as intermediaries between the research team and the community.

> **"We always work with NGOs because they are essential for connecting us with contacts, but also for advising us on how to design the surveys... you simply cannot do this without the participation of the populations you are sampling. They really need to be involved at every stage of the process."**
>
> Anonymized testimony. Interview series conducted for this publication.

PRIORITIZE TRAINING AND AWARENESS FOR THE FIELD TEAM

Interviewers and field supervisors should receive training not only on the use of the research instruments but also on the social and cultural characteristics of the population being studied. Such training helps establish consistent practices, promotes the use of respectful language, and improves interactions with participants.

> "When working with hidden populations, the field team must be sensitized to—or have either direct or indirect experience with—these populations. It is very difficult for someone with no connection to this context to effectively implement the study or accurately convey what the instrument is asking."
>
> Anonymized testimony. Interview series conducted for this publication.

CONSIDER TEAM SAFETY AND ACCESS TO THE FIELD

In some settings, gaining access to certain groups or geographic areas requires local intermediaries or informal permission. Coordination with community stakeholders helps ensure safe conditions for conducting fieldwork.

> "The same is true when working with gangs. [...] The only reason we were able to enter was because we had permission to come and go. You usually need a local intermediary to facilitate that process."
>
> Anonymized testimony. Interview series conducted for this publication.

PROVIDE SUPPORT AND DEBRIEFING OPPORTUNITIES FOR THE RESEARCH TEAM

Interviews addressing experiences of violence, discrimination, or other sensitive issues can place an emotional burden on field researchers. For this reason, some studies incorporate supervision or support sessions to help research staff process these experiences.

> "We had a supervision session after every meeting [...] where we talked about how we felt about the questions, how the pilot testing went, and what the experience brought up for us."
>
> Anonymized testimony. Interview series conducted for this publication.

RECOGNIZE AND FAIRLY COMPENSATE THE CONTRIBUTIONS OF RESEARCH STAFF AND PARTICIPANTS

Providing training and fair compensation to field staff—especially when the team includes members of the study population—helps strengthen their engagement and commitment to the research process. In addition, some studies use participant incentives, such as gift cards or monetary compensation, which can facilitate participation when implemented in a transparent and ethical manner.

> **"Not only were they given jobs, but they were also trained and compensated. That was extremely valuable because they felt they were part of the process."**
>
> Anonymized testimony. Interview series conducted for this publication.

Taken together, the interviews indicate that the composition of the field team is a critical methodological decision that should be made during the research design phase. A well-trained, diverse team with close ties to the study population can substantially improve field access, foster participants' trust, and enhance the quality of the data collected.

4. Selection of Methods

The interviews indicate that the selection of methods for studying hidden populations typically involves a combination of technical, contextual, and ethical considerations. In the absence of formal sampling frames, research teams adopt flexible strategies that facilitate access to the target population's social networks, foster trust, and ensure feasible operating conditions for fieldwork. Based on these experiences, the following recommendations emerged.

COMBINE METHODOLOGICAL APPROACHES TO FACILITATE FIELD ACCESS

Studies often combine convenience-based techniques with network-based approaches—such as Snowball Sampling, Respondent-Driven Sampling (RDS), and Time-Location Sampling (TLS)—to progressively gain access to the target population through intermediaries or trusted contacts.

> "For the identification and recruitment phase, we use techniques such as Snowball Sampling, where one person refers us to another because there is no formal sampling frame. In some cases, we work with collective key informants who already have the trust of the population, which ultimately opens doors that initially seemed impossible to access."
>
> Anonymized testimony. Interview series conducted for this publication.

COMPLEMENT QUANTITATIVE METHODS WITH QUALITATIVE APPROACHES

The interviews highlight the value of integrating qualitative methods—such as semi-structured interviews and focus groups—to better understand the internal dynamics, meanings, and forms of organization within the communities being studied. These approaches facilitate a gradual entry into the field and enrich the interpretation of quantitative findings.

> "We generally like to use focus groups to observe interactions among participants, and semi-structured interviews for key informants."
>
> Anonymized testimony. Interview series conducted for this publication.

> "Studies can include both qualitative interviews and surveys. For example, in one study, a survey questionnaire was developed and then complemented with a qualitative study."
>
> Anonymized testimony. Interview series conducted for this publication.

ADAPT SAMPLING STRATEGIES TO THE INSTITUTIONAL CONTEXT

The interviews indicate that data collection strategies also depend on the type of institution conducting the study. While some research teams rely on targeted, network-based methods, national statistical offices tend to favor broad-coverage approaches, such as open surveys disseminated through public awareness campaigns or partnerships with civil society organizations.

> **"In most cases, we do not use Snowball Sampling. Instead, we conduct an open survey available to the general population, promoted through widespread media campaigns and also through civil society organizations."**
>
> Anonymized testimony. Interview series conducted for this publication.

CONSIDER THE LOGISTICAL AND TECHNOLOGICAL CONDITIONS OF THE RESEARCH SETTING

Methodological decisions are also shaped by logistical and technological factors, such as digital connectivity, geographic accessibility, and the resources available for fieldwork. These constraints can directly influence how the target population is identified and the cost of data collection.

> **"[...] I remember we wanted to identify mothers with young children, so we sent interviewers into the field to create a list of households that met that criterion. They then returned to conduct a longer survey. It is a very costly process, but it allows you to obtain a representative sample."**
>
> Anonymized testimony. Interview series conducted for this publication.

BALANCE STATISTICAL REPRESENTATIVENESS WITH OPERATIONAL FEASIBILITY

In practice, method selection reflects an ongoing tension between the ideal of statistically robust designs—such as probability sampling—and the practical constraints of conducting fieldwork with hidden populations. As a result, methodological decisions often balance statistical rigor with pragmatic considerations, including community trust, the sensitivity of the topic under study, and the cultural appropriateness of the research instruments.

Taken together, the interviews indicate that **no single method is appropriate for all hidden populations**. The most effective strategies tend to combine multiple methodological approaches and adapt them to the social, institutional, and geographic contexts in which the research is conducted.

5. Ethics in Research

The interviews highlight that ethics is a cross-cutting principle in research involving hidden populations. Beyond compliance with formal protocols, ethical research practices require safeguarding the safety, confidentiality, and well-being of participants throughout the entire research process. Based on the experiences shared by the interviewees, the following recommendations emerged.

INCORPORATE ETHICS FROM THE STUDY DESIGN STAGE

Ethical considerations should be integrated from the earliest stages of the research process, including the design of research instruments, fieldwork planning, and the management of collected data. This requires anticipating potential risks and establishing mechanisms to protect the identity and safety of participants.

> **"Working with these types of populations not only requires flexible methods, but also an ethical commitment to ensuring confidentiality, safety, and the use of appropriate language."**
>
> Anonymized testimony. Interview series conducted for this publication.

COMBINE INTERNATIONAL STANDARDS WITH CONTEXT-SPECIFIC ADAPTATIONS

The interviews indicate that many studies adopt international ethical frameworks, which are then adapted to the local context with the support of community stakeholders and subject-matter experts.

> **"We always try to follow [...] best practices, but [...] there is nothing fundamentally different from international standards. What we do try to do is adapt those guidelines based on the advice we receive from the key individuals we involve in our working group when conducting research with hidden populations."**
>
> Anonymized testimony. Interview series conducted for this publication.

ENSURE CONFIDENTIALITY AND DATA ANONYMIZATION

Commonly used strategies include avoiding the collection of personally identifiable information, anonymizing datasets before analysis, and restricting access to sensitive data. In some cases, these measures are reinforced through institutional agreements or contractual provisions governing data use. In national statistical offices, such practices are often part of their standard institutional procedures.

> "The National Statistics Institute already has this data. There are no additional data protection measures because confidentiality is already built into all the information that is collected, since it includes sensitive information [...]."
>
> Anonymized testimony. Interview series conducted for this publication.

IMPLEMENT PRACTICES TO PROTECT PARTICIPANT SAFETY

In studies addressing sensitive topics—such as gender identity, sexual orientation, or violence—ethical practices also include operational decisions, such as selecting safe and neutral locations for interviews or using tools that ensure participants can provide responses privately.

> "In some regions, LGBT organizations allowed us to use their facilities. In others, we had to find a very neutral location because otherwise people would not feel comfortable entering."
>
> Anonymized testimony. Interview series conducted for this publication.

USE TECHNOLOGIES THAT ENHANCE PRIVACY DURING DATA COLLECTION

Some surveys incorporate self-administered devices or survey modules that allow participants to answer sensitive questions without interviewers observing their responses.

> "When participants reach the sections on gender identity and sexual orientation, they are given headphones [...] They interact with the device and the audio, and the interviewer neither sees nor hears what they are answering."
>
> Anonymized testimony. Interview series conducted for this publication.

In summary, the interviews indicate that ethics in research involving hard-to-reach populations extends beyond compliance with formal protocols. It requires a sensitive and context-specific approach that ensures participants' trust and safety throughout the research process.

6. Data Quality

The interviews highlight that the quality of data collected in studies involving hidden populations depends not only on methodological rigor but also on relational and operational factors, such as trust within the community, the clarity of research instruments, and the cultural sensitivity of the field team. In this regard, the interviewees identified the following key recommendations.

DESIGN CLEAR AND CULTURALLY APPROPRIATE RESEARCH INSTRUMENTS

Questionnaires should avoid technical jargon and use language that is clear and easily understood by the target population. This facilitates participation and reduces the likelihood of misinterpretation during data collection.

> "When we present the questions, or when a respondent asks the interviewer, 'What does this mean?' or 'How should I understand this?', the interviewer needs to be able to explain it or find simpler, more accessible wording so that the respondent feels comfortable."
>
> Anonymized testimony. Interview series conducted for this publication.

TRAIN THE FIELD TEAM TO UNDERSTAND CULTURAL NORMS AND INTERNAL COMMUNITY DYNAMICS

The interviewees noted that the populations being studied often have their own internal norms and symbolic hierarchies, which can influence both participants' willingness to take part in the study and the quality of the responses obtained.

> "These intragroup codes are difficult to navigate when trying to collect reliable information [...] if you get them wrong [...] you can completely undermine the data collection process."
>
> Anonymized testimony. Interview series conducted for this publication.

MAINTAIN ANALYTICAL CONSISTENCY IN THE CATEGORIES USED

In studies addressing gender identity, sexual orientation, or other social characteristics, it is important to use clear and consistent categories to ensure that findings are interpreted appropriately and can be compared over time.

IMPLEMENT MONITORING MECHANISMS THROUGHOUT FIELDWORK

For network-based sampling methods, such as Respondent-Driven Sampling (RDS), study quality depends on monitoring recruitment chains, identifying potential bottlenecks, and verifying sample convergence.

> "In addition to avoiding bottlenecks, it is essential to ensure that the sample reaches convergence [...] by monitoring long recruitment chains."
>
> Anonymized testimony. Interview series conducted for this publication.

ESTABLISH ONGOING VERIFICATION AND FEEDBACK PROCESSES

Verifying data through follow-up interviews, maintaining regular communication with field teams, and periodically reviewing the information collected help identify inconsistencies and correct errors during the data collection process.

> "Maintaining contact with the people who were interviewed to verify that they were, in fact, interviewed."
>
> Anonymized testimony. Interview series conducted for this publication.

Taken together, these findings indicate that the quality of data collected in studies involving hidden populations depends not only on the rigor of the statistical methods employed, but also on building trusting relationships and ensuring thoughtful, respectful interactions with participants.

7. Public Policy Impact

The interview findings indicate that generating high-quality data on hidden populations has both instrumental and transformative value. Such evidence not only brings legitimacy to historically invisible realities but also supports public decision-making and strengthens policy evaluation processes.

> **"Producing evidence on hidden populations is not merely a technical challenge—it is an ethical commitment."**
>
> Anonymized testimony. Interview series conducted for this publication.

Access to disaggregated and reliable data makes it possible to identify socioeconomic disparities and patterns of exclusion, thereby informing the design and improvement of public policy and strengthening evidence-based advocacy in policy discussions.

> **"Having reliable data for any policy design [...] means that your arguments will carry much more weight than arguments based solely on advocacy or activism without supporting data."**
>
> Anonymized testimony. Interview series conducted for this publication.

Moreover, statistical visibility has symbolic and political value. It empowers communities by demonstrating that their experiences and needs are recognized, thereby legitimizing their demands and strengthening their participation in the public sphere.

> **"It not only generates solid evidence for institutions, but also empowers the population by showing that their experiences have social and political value—value that has historically been denied to them."**
>
> Anonymized testimony. Interview series conducted for this publication.

Taken together, these findings show that generating data on hidden populations serves not only to measure their characteristics but also to make inequalities visible, legitimize social claims, and inform policies with meaningful impact. Evidence thus becomes a key tool for developing more inclusive and context-sensitive public policies.

8. Challenges and Limitations

The interviews also identified several barriers that complicate research involving hidden populations. These challenges can be grouped into five main categories: access to the target population, methodological challenges, logistical constraints, institutional barriers, and challenges related to coordination with community stakeholders.

BARRIERS TO ACCESSING HIDDEN POPULATIONS

Distrust and fear of exposure are common barriers to participation, particularly among individuals who have experienced discrimination or stigma, such as LGBTQ+ populations and migrants. Participation may also be affected by survey fatigue resulting from repeated research efforts or by expectations related to financial incentives.

METHODOLOGICAL CHALLENGES

The absence of appropriate sampling frames makes it difficult to obtain representative samples and adequately capture the internal diversity of the populations being studied. As a result, certain groups—such as transgender individuals, Indigenous peoples, and lesbian women—may be underrepresented.

In addition, the inappropriate use of specialized techniques—such as Respondent-Driven Sampling (RDS) or Time-Location Sampling (TLS)—can compromise the quality of findings when their methodological assumptions are not met or when they are mistakenly interpreted as producing fully representative samples.

LOGISTICAL AND RESOURCES CONSTRAINTS

Limited funding, tight implementation timelines, and technological gaps—such as poor connectivity in remote areas—can restrict the scope of fieldwork and reduce the feasibility of implementing more complex methodological approaches.

INSTITUTIONAL AND REGULATORY BARRIERS

In some countries, regulatory frameworks do not permit the collection or publication of information on certain social characteristics, such as race, ethnicity, or sexual diversity. These constraints are often compounded by the heavy workloads of many public institutions, which limit their capacity to undertake additional research initiatives.

> "National Statistical Offices face limitations in dedicating time to additional initiatives because of their heavy workloads [...] This is compounded by the national regulatory framework, which in many cases neither requires nor permits the collection of information on sensitive topics such as race, ethnicity, or LGBTQ+ populations, thereby limiting progress even when there is institutional interest."
>
> Anonymized testimony. Interview series conducted for this publication.

CHALLENGES IN INTERINSTITUTIONAL AND COMMUNITY COORDINATION

Access to hidden populations often depends on collaboration with civil society organizations, community leaders, or other intermediaries. While these stakeholders play a critical role in building trust and facilitating fieldwork, challenges may arise because of differing expectations, concerns about how the information will be used, or participation fatigue resulting from repeated involvement in research studies.

In addition, in some settings, access to certain communities requires specific intermediaries or informal permission, which may further constrain the scope of the research.

Table 5. Main Challenges Reported by Interviewees in Researching Hidden Populations, by Category

Category	Subcategory	Details
Barriers to Accessing Hidden Populations	Distrust	Fear of exposure or discrimination has been reported, particularly among LGBTQ+ populations and migrants, due to prior experiences of stigma.
	Survey fatigue or lack of interest	Participants may experience fatigue when they are repeatedly invited to take part in research activities.
	Conditional participation	Participation may, in some cases, depend on the availability and amount of financial incentives.[14]
Methodological Challenges	Representativeness	Lack of appropriate sampling frames, particularly for capturing the internal diversity of the population, resulting in the underrepresentation of certain groups (e.g., transgender individuals, Indigenous peoples, or lesbian women).
	Application of advanced sampling methods	Inappropriate use of methods such as Respondent-Driven Sampling (RDS) or Time-Location Sampling (TLS) without meeting their underlying methodological assumptions.
	Conceptual misunderstanding	Use of non-probability sampling techniques that are subsequently misinterpreted as producing representative samples.
	Continuity of analysis and longitudinal perspective	Once a sample has been collected for a given year, it may be difficult to replicate the research instrument or maintain comparability over time because of the evolving nature of social networks.
Logistical and Resource Constraints	Insufficient resources	Budget limitations that may restrict the scope or depth of the study.
	Technological and geographic gaps	Difficulties in surveying individuals in areas with limited connectivity or in remote locations.
	Technical capacity	Limited expertise in specialized research methods and in the ethical management of sensitive information.
Institutional and Regulatory Barriers	Regulatory frameworks	Regulatory barriers may limit access to information, hinder the implementation of more rigorous methodological approaches, or restrict the collection of broader and more disaggregated data.
Interinstitutional and Community Coordination	Dependence on local intermediaries	Reliance on intermediaries to gain access to communities or closed groups, such as gangs.
	Tensions between the research team and local organizations	Differences in expectations, distrust of the study, or conflicts related to the use of the information collected.

Source: Prepared by the authors.

14. This condition has a structural basis: hidden populations often share characteristics such as exclusion from the formal labor market or experiences of underemployment. This requires a careful balance. On the one hand, individuals may be reluctant to participate unless they receive financial compensation they consider adequate, particularly when they are in precarious employment situations. On the other hand, research teams must ensure that incentives do not create inequities or contribute to conditions of underemployment.

Conclusions

Research involving hard-to-reach populations requires recognizing that certain groups face structural barriers that limit their social and economic inclusion. These barriers create information gaps that affect the quality of public policy design, implementation, and evaluation, as well as the effectiveness of interventions.

The statistical invisibility of these groups has direct implications for resource allocation, the effectiveness of programs, and the capacity to promote inclusive growth. Adopting appropriate approaches to identify and measure these populations strengthens evidence-based decision-making, improves program targeting, and enhances the impact of interventions on people's quality of life.

There is no universally applicable method for researching hidden populations. The choice of strategy depends on the context, the research objectives, and the characteristics of the population being studied. Among the most commonly used methods are:

- **Time-Location Sampling (TLS),** recommended when there are clearly identifiable venues where the target population regularly gathers;
- **Targeted Sampling (TS),** useful for geographically concentrated populations and when the objective is to study specific subgroups;
- **Respondent-Driven Sampling (RDS),** effective when social networks are dense and well connected, provided that the number of referrals is controlled and network size is documented; and
- **Snowball Sampling,** appropriate during exploratory phases, when the primary goal is to generate descriptive insights rather than support statistical inference.

Complementing these approaches with qualitative techniques, such as **Mystery Shopping** and **Vignette Studies**, makes it possible to capture nuances, better understand internal dynamics, and make informed adjustments during fieldwork.

Operationally, the involvement of members of the study population and the development of trusting relationships with community leaders and other key stakeholders are essential. These strategies facilitate access, improve data quality, and reduce barriers to participation. They also help ensure safe environments for data collection, support the application of appropriate anonymization techniques, and promote the use of respectful, culturally appropriate language.

Research involving hidden populations also requires navigating the tension between what can be measured and what remains invisible, balancing the need to generate evidence with the ethical obligation to avoid causing further harm to these populations. No single method can address all the technical, logistical, or ethical challenges involved. The findings show that the most effective strategies combine methodological pluralism, contextual sensitivity, community participation, and a strong ethical commitment.

Finally, examining these populations through a rigorous methodological lens is an act of public responsibility. Producing data is not only about measurement—it is also about making inequalities visible, recognizing the experiences of groups that have historically remained invisible, and informing the design of impactful public policies, thereby strengthening both the technical evidence base and the socioeconomic inclusion and quality of life of these populations.

References

Allen, S. T., Footer, K. H. A., Galai, N., et al. (2019). Implementing targeted sampling: Lessons learned from recruiting female sex workers in Baltimore, Maryland. Journal of Urban Health. https://doi.org/10.1007/s11524-018-0292-0

Allison, P., Severt, D., and Dickson, D. (2010). A conceptual model for mystery shopping motivations. Journal of Hospitality Marketing and Management. https://doi.org/10.1080/19368623.2010.493077

Atkinson, R., and Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. Social Research Update, University of Surrey.

Bauermeister, J. A., Pingel, E. S., Jadwin-Cakmak, L., Meanley, S., Alapati, D., Moore, M., Lowther, M., Wade, R., and Harper, G. W. (2015). The use of mystery shopping for quality assurance evaluations of HIV/STI testing sites offering services to young gay and bisexual men. AIDS and Behavior, 19(10), 1919–1927. https://doi.org/10.1007/s10461-015-1174-z

Chao, A. (1998). Capture–recapture. In P. Armitage & T. Colton (Eds.), Encyclopedia of biostatistics (pp. 482–486). Wiley.

Chao, A. (2001). An overview of closed capture–recapture models. Journal of Agricultural, Biological, and Environmental Statistics, 6(2), 158–175.

Cheah, C. W., Koay, K. Y., and Xiang, K. (2023). Uncovering the voices of marginalized minorities in public policy research: A critical review of the image and text-based vignette method. International Journal of Qualitative Methods. https://doi.org/10.1177/16094069231163429

Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.

Coleman, J. S. (1958). Relational analysis: The study of social organizations with survey methods. Human Organization. http://www.jstor.org/stable/44124097

De Castro, F., et al. (2018). Adolescent access to information on contraceptives: A mystery client study in Mexico. Journal of Adolescent Health. https://doi.org/10.1016/j.jadohealth.2017.08.001

Dusek, G. A., Yurova, Y. V., and Ruppel, C. P. (2015). Using social media and targeted snowball sampling to survey a hard-to-reach population: A case study. International Journal of Doctoral Studies. http://ijds.org/Volume10/IJDSv10p279-299Dusek0717.pdf

Equilibrium Business, Data and Communities (BDC). (2025). Research for the Strengthening of the LGBTIQ+ Business and Labor Fabric in Panama. https://equilibriumbdc.com/publicaciones/investigacion-para-el-fortalecimiento-del-tejido-empresarial-y-laboral-lgbtiq-en-panama/

Espinosa, M. P., and Brañas-Garza, P. (Eds.). (2024). Experimental Methods in Economics. Editorial Centro de Estudios Ramón Areces.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. Social Problems. https://doi.org/10.2307/3096941

Heckathorn, D. D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. Social Problems. https://doi.org/10.1525/sp.2002.49.1.11

Inter-American Development Bank. (2023). IDB Impact+: Marco de gestión de impacto para el Desarrollo [*Impact Management Framework for Development*]. https://www.iadb.org/es/inicio/bidimpact

IDB Group Institutional Strategy 2024–2030: Transformación para una mayor escala e impacto [*Transformation for Greater Scale and Impact*]. https://www.idbinvest.org/sites/default/files/2024-03/Grupo-BID-Estrategia-Institucional-2030.pdf

International Organization for Migration. (2020). Glossary on Migration (IML No. 34). International Organization for Migration.

Karon, J. M., and Wejnert, C. (2012). Statistical methods for the analysis of time–location sampling data. Journal of Urban Health. https://doi.org/10.1007/s11524-012-9676-8

McInroy, L. B., and Beer, O. W. J. (2022). Adapting vignettes for internet-based research: Eliciting realistic responses to the digital milieu. International Journal of Social Research Methodology. https://doi.org/10.1080/13645579.2021.1901440

Morena de Diago, B. (2013). Mystery shopping: Indicadores de estudio en bibliotecas. Revista General de Información y Documentación, 23(2), 369–386.

Mosler Vidal, E. (2021). No dejar a ningún migrante atrás: La Agenda 2030 y el desglose de datos [*Leave No Migrant Behind: The 2030 Agenda and Data Disaggregation*]. International Organization for Migration.

Paz-Bailey, G., et al. (2013). Reaching men who have sex with men: A comparison of respondent-driven sampling and time-location sampling in Guatemala City. AIDS and Behavior. https://doi.org/10.1007/s10461-013-0589-7

Poorolajal, J., Mohammadi, Y., and Farzinara, F. (2017). Using the capture–recapture method to estimate the human immunodeficiency virus-positive population. Epidemiology and Health. https://doi.org/10.4178/epih.e2017042

Raifman, S., et al. (2022). Respondent-driven sampling: A sampling method for hard-to-reach populations and beyond. Current Epidemiology Reports. https://doi.org/10.1007/s40471-022-00287-8

Rozo, S. (2021). Tips for collecting surveys of hard-to-reach populations. World Bank Blogs. https://blogs.worldbank.org/impactevaluations/tips-collecting-surveys-hard-reach-populations

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2023). Research methods for business students (9th ed.). Pearson.

Semaan, S., Lauby, J., & Liebman, J. (2002). Street and network sampling in evaluation studies of HIV risk-reduction interventions. AIDS Reviews, 4(4), 213–223.

Shaghaghi, A., Bhopal, R. S., and Sheikh, A. (2011). Approaches to recruiting hard-to-reach populations into research: A review of the literature. Journal of Health Promotion Perspectives. https://hpp.tbzmed.ac.ir/Article/HPP_73_20120617105704

Soltanian, A. R., et al. (2020). Estimation of the hidden population with COVID-19 disease. International Journal of Maternal and Child Health and AIDS. https://doi.org/10.21106/ijma.396

Ting, H., et al. (2025). Snowball sampling: A review and guidelines for survey research. Asian Journal of Business Research. https://doi.org/10.14707/ajbr.250186

United Nations Sustainable Development Group. (2022). Operationalizing leaving no one behind: Good practice note for UN country teams.

Watters, J. K., and Biernacki, P. (1989). Targeted sampling: Options for the study of hidden populations. Social Problems. https://doi.org/10.2307/800824

World Health Organization, and Joint United Nations Programme on HIV/AIDS. (2010). Guidelines on surveillance among populations most at risk for HIV.

World Health Organization, and United Nations Human Settlements Programme. (2010). Hidden cities: Unmasking and overcoming health inequities in urban settings. https://iris.who.int/handle/10665/44439

Yauck, M., Moodie, E. E. M., Apelian, H., Peet, M.-M., Lambert, G., Grace, D., Lachowsky, N. J., Hart, T., & Cox, J. (2020). Sampling from networks: Respondent-driven sampling. McGill University.

Annex

List of Reviewed Studies on Hidden Populations

Implementing Organization	Funding Organization	Year	Country	Title and Direct Access	Method
IPA Colombia (Innovations for Poverty Action)	Inter-American Development Bank (BID/IDB), World Bank Research Support Budget grant, USC Marshall School of Business	2022	Colombia	Life Out of the Shadows: Impacts of Amnesties in the Lives of Migrants	Combines stratified sampling (Administrative Registry of Venezuelan Migrants, RAMV, group) with snowball/referrals (non-RAMV group).
IDB	IDB	2020	Asia and other continents	Toward a Better Understanding of Discrimination Based on Sexual Orientation and Gender Identity	Studies using various sampling methods, including Respondent-Driven Sampling (RDS) and Snowball Sampling; probability and nonprobability sampling designs
CDC, Universidad del Valle de Guatemala	CDC	2013	Guatemala	Reaching Men Who Have Sex with Men: A Comparison of Respondent-Driven Sampling and Time-Location Sampling in Guatemala City	Time-Location Sampling y Respondent-Driven Sampling
DANE	DANE	2022	Colombia	Mercado laboral de la población LGBT	Probability, multistage, stratified, and cluster sampling
Equilibrium, International Organization for Migration (IOM)	IOM	2024	Peru	Análisis exploratorio de la contribución fiscal y económica de la migración venezolana en Perú	Snowball Sampling
Equilibrium		2025	Panama	Investigación para el Fortalecimiento del Tejido Empresarial y Laboral LGBTIQ+ en Panamá	Respondent-Driven Sampling
INEI-ENPOVE	INEI	2023	Peru	II Encuesta a la Población Venezolana residente en el Perú	Stratified sampling
INEI-LGBTI	INEI	2017	Peru	Primera Encuesta Virtual para Personas LGBTI, 2017	Exploratory, nonprobability sampling
Nova Southeastern University	Nova Southeastern University	2015	United States and Russia	Using Social Media and Targeted Snowball Sampling to Survey a Hard-to-Reach Population: A Case Study	Snowball Sampling and Targeted Sampling (TS)
World Health Organization (WHO)	World Health Organization (WHO)	2010	Pakistan	Unmasking and Overcoming Health Inequities in Urban Settings	Targeted Sampling (TS)
World Health Organization (WHO), Joint United Nations Programme on HIV/AIDS (UNAIDS)	WHO	2011	China and other countries	Guidelines on surveillance among populations most at risk for HIV	Collection of studies including Respondent-Driven Sampling (RDS)
UNFPA	UNFPA	2017	Dominican Republic	Población inmigrante venezolana	Additional phase using Snowball Sampling and a national probability sample

Implementing Organization	Funding Organization	Year	Country	Title and Direct Access	Method
UNICEF	UNICEF	2020	Southern Route from Eastern Africa	An unsettled past, an uncertain future	Snowball Sampling
-	U.S. National Institute of Mental Health (NIMH), through the HIV Center for Clinical and Behavioral Studies at the New York State Psychiatric Institute and Columbia University	2011	Argentina	Use of Respondent Driven Sampling (RDS) Generates a Very Diverse Sample of Men Who Have Sex with Men (MSM) in Buenos Aires, Argentina | PLOS One	Respondent-Driven Sampling
IDB	IDB	2024	Eight Latin American countries	Socio-Economic Disparities in Latin America among Same-Sex and Different-Sex Couples (Technical Note)	National statistics (census data)
IDB	IDB	2025	Mexico	The Lives of Intersex People: Socio-Economic and Health Disparities in Mexico	National statistics (national survey)
IDB	IDB	2023	Ecuador	Measuring Labor Market Discrimination against LGTBQ+ in the Case of Ecuador: A Field Experiment	Respondent-Driven Sampling
IDB	IDB	2023	Argentina, Colombia, Ecuador, and Peru	Discrimination Against Gay and Transgender People in Latin America: A Correspondence Study in the Rental Housing Market	Correspondence study
IDB	IDB	2023	Mexico	Challenges for Measuring the LGBT+ Population and Homophobia in Mexico	Item count technique (ICT)
IDB	IDB	2023	Colombia	Is there discrimination against children of same-sex households? Evidence from an experimental study in Colombia	Correspondence study
IDB	IDB	2023	Colombia	How Accurately are Household Surveys Measuring the Size and Inequalities for the LGBT Population in Bogota, Colombia? Evidence from a List Experiment	National statistics (household survey)
IDB	IDB	2022	Uruguay	Does Gender and Sexual Diversity Lead to Greater Conflict in the School?	Data from 95 schools and secondary schools
IDB	IDB	2024	Latin America	The Migrant Penalty in Latin America: Experimental Evidence from Job Recruiters	LinkedIn job advertisements and Respondent-Driven Sampling (RDS)


IDB
Improving lives