

**Proceedings of Statistics Canada Symposium 2022:
Data Disaggregation: building a more representative data portrait of society**

Evaluating sampling methods for ethnic minorities

by Mariel McKone Leonard

Release date: March 25, 2024



Evaluating Sampling Methods for Ethnic Minorities

Mariel McKone Leonard¹

Abstract

Ethnic minorities are often underrepresented in survey research, due to the challenges many researchers face in including these populations. While some studies discuss several methods in comparison, few have directly compared these methods empirically, leaving researchers seeking to include ethnic minorities in their studies unsure of their best options.

In this article, I briefly review the methodological and ethical reasons for increasing ethnic minority representation in social science research, as well as challenges of doing so. I then present findings from ten studies which empirically compare methods of sampling and/or recruiting ethnic minority individuals. Finally, I discuss some implications for future research.

Key Words: Ethnic minorities, Hard-to-survey populations, Sampling.

1. Introduction

1.1 Importance of including ethnic minorities

Inclusion of demographic subgroups in research, either within studies focused on representative samples of these subgroups, or through larger subsamples within population surveys, is necessary both for improving survey data quality (Willis et al., 2014) and addressing human rights concerns (European Commission, 2021). In particular, the collection of data disaggregated based on racial or ethnic identity is essential to the production of statistics that support equality and anti-discrimination policies (European Commission, 2021). However, individuals with a migration background or who are members of ethnic minority groups are frequently underrepresented in survey research, even though members of these demographic subgroups may – in aggregate – comprise a sizeable percentage of the population.² Despite the importance of achieving diversity, inclusion, and equity in social science research, the field as a whole has struggled to incorporate ethnic minority populations and other demographic subgroups. Instead, many researchers have quietly accepted the exclusion of these groups from social science research, or at best, relied on “ad hoc convenience samples” that Sudman and Kalton (1986, p. 402) describe as “totally inadequate.” In the next section, I address the challenges of including ethnic minorities in social science research.

1.2 Challenges in including ethnic minorities

Demographic subgroups, have long been acknowledged by survey methodologists and statisticians as being difficult to include in surveys for numerous reasons, including sampling, linguistic, and access barriers which strain limited resources. In conjunction with rising research costs and falling participation and response rates, incorporating demographic subgroups into social science research has become increasingly challenging. While these challenges are not limited to ethnic minorities (Tourangeau et al., 2014), ethnic minority individuals may be particularly “hard-to-survey”. Willis et al. (2014) categorize hard-to-survey persons and groups as (1) hard-to-select, including hard-to-sample and hard-to-identify; (2) hard-to-recruit, including hard-to-locate and hard-to-persuade; and (3) hard-to-

¹Mariel McKone Leonard, Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM), Mauerstraße 76, 10117 Berlin, Germany.

² For example, in Germany, while no ethnic minority group is larger than 10 percent of the population, all together, individuals with a “migration background” comprise approximately one-quarter of the total German population (Destatis, 2022).

interview. Although multiple categorizations are possible for many demographic subgroups, ethnic minorities in particular appear more likely to present intersectional challenges.

Many countries – even those that classify individuals based on racial or ethnic identity – may not have available sampling frames for ethnic minority individuals. Where racial or ethnic identity is available for census tracts, or ethnic minority individuals cluster in specific geographic locations, stratified or clustered sampling methods, or time-location sampling, may be possible. Countries that maintain population registries, such as many European countries, often do not include – or even collect – racial or ethnic identity information in their registries, for reasons of data protection (Simon, 2012). Additionally, many countries do not clearly define terms related to racial and ethnic identity (Aspinall, 2002; Connelly, et al., 2016). While English-speaking countries often ask individuals to self-identify according to defined racial or ethnic groups, this is less common in continental Europe. Instead, European countries often rely on categories such as “immigrants” or those with a “migration background” (Simon, 2012). These terms may not resonate with individuals without direct migration experience, i.e., second or subsequent generation individuals, and may leave white-appearing ethnic minorities who experience discrimination – such as Jewish and Roma peoples – in a methodological blind spot (Aspinall 2002).

Even when groups of interest may be clearly defined, individual members of these ethnic groups may be challenging to find. Many ethnic minority groups are what Kalton and Anderson (1986) consider “rare”, i.e., they comprise less than 10 percent of the total population. As most probability-based sampling methods are resource intensive, including ethnic minorities in studies of the general population is often not cost effective (Nam et al., 2013). Some ethnic minority groups, such as Roma/Sinti, and Bedouin, are also highly mobile, making them additionally challenging to locate.

Finally, ethnic minority individuals are often systematically and significantly overrepresented among the most vulnerable or deprived members of society. They may face linguistic difficulties, particularly if they are also immigrants or excluded from educational opportunities; they may be unwilling to participate in research or speak to strangers due to historic experiences of discrimination (McDougal et al., 2001; Lynn et al., 2018).

Despite these difficulties, an ever-expanding body of literature discusses and advances methods of including hard-to-survey groups into social science research (e.g., Marpsat and Razafindratsima, 2010; Reichel and Morales, 2017). The bulk of this research, appears to focus on proposing and developing sampling methods for improving representation of demographic subgroups, including ethnic minorities. While these methods are essential to addressing the challenges of improving representation of hard-to-survey groups, they address a wide-range of demographic subgroups, not exclusively ethnic minorities. Moreover, only a few studies have directly compared these methods empirically. As a result, researchers seeking to improve representation of ethnic minorities in their studies may be overwhelmed by the methodological possibilities and unsure how best to evaluate their options.

2. The Present Study

2.1 Extant empirical comparisons of methods

Motivated by the lack of clear guidance for determining which sampling and recruiting methods are appropriate under what circumstances, I conducted a review of the related literature. I wanted specifically to understand how different methods compare in their effectiveness. I therefore focus on studies which (1) directly and empirically compare sampling and recruiting methods of (2) ethnic minority populations. Based on searches on Web of Science and Google Scholar, as well as reviewing citations of relevant studies, I was able to identify nine papers which meet this criteria. I additionally include findings from an unpublished study I conducted (Leonard, 2021). Table 2.1-1 presents these results.

Table 2.1-1
Studies providing empirical comparisons of methods of sampling ethnic minorities

Study	Method	Target Population	Study Location	Findings	Study Costs
Brand et al. (2019)	Registry; CBO	People of Turkish origin	Essen, Berlin, Heidelberg, Germany	Registry had lower response rate but less bias; CBO over-represented women and most marginalized	Not given
Cabral et al. (2003)	Registry; RDD; CBO	African; Latino Americans	San Francisco, United States	All methods over-represented lower socio-economic status; CBOs had highest eligibility and enrolment rates	Not given
Harris et al. (2003)	Health organizations; snowball; media	African American smokers	Midwestern United States	Health organizations over-recruited socio-economic status disadvantaged, networks more successful than health organizations. Media ads most successful	Health organizations + networks (159 USD); media (22 USD)
Hughes et al. (1995)	Registry; CBO; snowball	Black Caribbean/ African; S. Asian; E. Asians	Bristol, England	Snowball produced most interviews, no discussion of sample variance by method	Not given
Leonard (2021)	Registry; RDS; internet; TLS	People of Portuguese origin	Germany	Only internet sample of adequate size for subgroup analysis; registry under-represented individuals with recent migration, lower German language skills	Registry (70 Euro); RDS (7 Euro); internet (2Euro); TLS (15 Euro)
Lindsay et al. (2020)	CBO; snowball; internet	Brazilian immigrant families	Massachusetts, Rhode Island, United States	No discussion of sample variance by method	Not given
McDougal et al. (2001)	RDD; snowball	White and African Americans aged 70+	Cleveland, United States	RDD produced insufficient sample of African Americans. Additional respondents recruited via snowball sampling. Snowball sampling over-represented those with close community ties	Not given
McKenzie and Mistiaen (2007)	SRS with household listing; snowball; TLS	Nikkei (Japanese-Brazilian) households	Sao Paulo, Parana, Brazil	TLS and snowball over-represented those with close community ties	SRS (212 USD); snowball (100 USD); TLS (30 USD)
Vu et al. (2021)	CBO; internet; snowball	People of Vietnamese origin	United States	All methods over-represented women, higher socio-economic status. CBOs had lowest overall response rate; internet groups had highest followed by snowball	Not given

**Table 2.1-1
Studies providing empirical comparisons of methods of sampling ethnic minorities**

Study	Method	Target Population	Study Location	Findings	Study Costs
Weinmann et al. (2019)	Registry; RDS	Syrian immigrants	Munich, Germany	RDD under-represented individuals with German citizenship, registry under-represented socially vulnerable individuals, including lower socio-economic status	Not given

2.2 Summary of findings

The majority of studies focus on the methodological options available for sampling ethnic minorities and thus increasing representation and inclusion of research studies. Therefore, these studies employ both nonprobability- and probability-based methods as available. The most commonly used nonprobability-based methods include referrals from community-based organizations (CBOs) which may include social clubs, community health clinics, or other similar organizations; nonprobability-based network methods (e.g., snowball sampling); and both traditional and non-traditional (e.g., social media, internet) media advertising. Only McKenzie and Mistiaen (2007) and Weinmann et al. (2019) focus on achieving representative samples of their target populations. As a result, these studies use only methods which are capable of yielding probability-based samples, including variations of random sampling (e.g., from population registry lists, random-digit-dialing (RDD), or stratified random sampling (SRS) of census tracts); respondent-driven sampling (RDS); or time-location sampling (TLS; also called center sampling). RDS is a network-based method which uses a link-tracing design to produce Markov chains to allow the researcher to determine the sampling probability of each respondent (Gile and Handcock, 2010). In TLS, researchers randomly sample locations identified as places where the target population frequents. Individuals are then sampled at each location, and based on their frequency of visits, a sampling probability is developed (Baio, et al., 2011).

Community-based methods seem to consistently increase inclusion of more vulnerable individuals, but also over-represent those with close community ties (McKenzie and Mistiaen, 2007; Brand et al., 2018). These findings are not inconsistent, as socially disadvantaged individuals may be more likely to depend on community or mutual aid (Hodges 2016; Zemore et al. 2021). Similarly, network-based methods such as snowball and RDS increase response rates and reach of vulnerable individuals (Hughes et al., 1995; Harris et al., 2003). However, both community-based methods and network-based methods require a sufficient number of individuals in a specific geographic area who are also active in community life (Brand et al., 2018). This may decrease these methods' utility for developing nation-wide studies, or among groups whose members are dispersed or not well-connected. Media-based recruitment and sampling methods may be more successful at producing geographically dispersed samples.

While all authors report success in improving inclusion of their target populations, no study claims to produce fully representative samples, and all identify significant biases resulting from either over- or under-representation of certain subsections of the target populations (see Table 2.1-1). Furthermore, although only three of the 10 studies (Harris et al., 2003; Leonard, 2021; McKenzie and Mistiaen, 2007) reviewed provided a breakdown of costs of by method, traditional probability-based sampling methods which require more intensive preparation, such as name-based pre-classification of lists or stratification, were consistently more expensive. Network-based sampling methods, which may allow for probability-based sampling, varied in their cost-effectiveness. Recruitment through media – social or traditional – appears to be the least expensive method, but this method is only able to produce nonprobability samples. For studies in which accessing the widest possible range of individuals, i.e., maximizing representation and inclusion, rather than statistical generalizability, is the primary goal, community-based methods or media recruiting may be most effective (Harris et al. 2003).

3. Conclusion

3.1 Implications for future research

While research based in the U.S., focused on non-rare racial and ethnic minority groups, is still the largest subcategory of literature, a growing number of studies have begun to target rare ethnic minority groups. Although the rarity of a group is not the only challenge researchers seeking to increase inclusion of minority individuals face, it is a significant one. Improving knowledge and understanding of how to target rare ethnic minority groups in particular will allow researchers in a growing number of countries and regions to improve the diversity of their surveys. However, improving representativity within social science research remains difficult. Even with the use of probability-based methods which are able to produce representative samples of majority populations, the studies included in this review were unable to produce unbiased samples of their target populations. Statistically “representative” samples derived from probability-based methods may not fully capture the diversity of a minority group, while nonprobability-based methods often result in increased “representation” (inclusion) of a more broadly diverse sample that cannot be considered representative. The gap between representativity and representation must be addressed to improve data quality and produce generalizable research results.

It remains an open question of how best to combine probability- and nonprobability-based samples, or even multiple nonprobability-based samples, should the use of nonprobability-based samples be necessary to produce more diverse samples. It is possible that using a multimethod approach as is common in multinational, multiregional, and multicultural (3MC) studies may produce more representative and generalizable results. Additionally, recent work by Lohr (2021) and Rao (2021) suggest ways in which multi-frame sample estimators may be adopted to improve weighting of nonprobability-based samples, particularly in the absence of adequate probability samples.

Finally, while the use of more efficient and effective sampling and recruiting does increase the likelihood of including ethnic minorities, addressing only the selection and recruitment of ethnic minorities does not solve challenges presented by communication difficulties or lack of trust in social science research. It is important that the field of survey methodology, and social science research more broadly, continue to conduct comparative studies of methods of including ethnic minorities and other demographic subgroups in research. Particularly, consideration of lessons learned, or guidelines for selecting and adapting methodological options will allow future researchers to improve inclusion of ethnic minorities and draw conclusions about best practices more reliably.

References

- American Association of Public Opinion Research [AAPOR], (n.d.), “AAPOR 77th Annual Conference”, Retrieved: <https://www.aapor.org/Conference-Events/Recent-Conferences/2022-Annual-Conference-Recap.aspx>
- Aspinall, P.J. (2002), “Collective Terminology to Describe the Minority Ethnic Population: The Persistence of Confusion and Ambiguity in Usage”, *Sociology*, 36(4), pp. 803–816.
- Baio, G., Blangiardo, G.C., and Blangiardo, M. (2011), “Centre Sampling Technique in Foreign Migration Surveys: A Methodological Note”, *Journal of Official Statistics*, 27(3), pp. 451-465.
- Bilsborrow, R.E., Hugo, G., Oberai, A.S. and Zlotnik, H. (1997), “International Migration Statistics: Guidelines for Improving Data Collection Systems”, unpublished report, Geneva, Switzerland: International Labour Office.
- Borges, M.J. (2011), “Portuguese Labor Migrants in Western and Central Europe since the 1950s: The Examples of France and Germany”, in Bade, K.J., et al. (eds) *The Encyclopaedia of Migration and Minorities in Europe*, Cambridge: Cambridge University Press, pp. 619-623.
- Brand, T., Samkange-Zeeb, F., Dragano, N., et al. (2018), “Participation of Turkish Migrants in an Epidemiological Study: Does the Recruitment Strategy Affect the Sample Characteristics”, *Journal of Immigrant and Minority Health*, 21(4), pp. 811-819.
- Cabral, D.N., Nápoles-Springer, A.M., Miike, R., et al. (2003)., “Population- and Community-Based Recruitment of African Americans and Latinos”, *American Journal of Epidemiology* 158(3), pp. 272-279.

- Connelly, R., Gayle, V., Lambert, P.S. (2016), "Ethnicity and Ethnic Group Measures in Social Survey Research", *Methodological Innovations Online*, 9, pp. 1-10.
- European Commission (2021), "Guidance Note on the Collection and Use of Equality Data Based on Racial or Ethnic Origin", unpublished report, Brussels, Belgium: European Commission.
- Gile, K.J., and Handcock, M.S. (2010), "Respondent-Driven Sampling: An Assessment of Current Methodology", *Sociological Methodology*, 7, pp. 285-327.
- Harris, K.J., Ahluwalia, J.S., Catley, D., et al. (2003), "Successful Recruitment of Minorities into Clinical Trials: The Kick It at Swope Project", *Nicotine & Tobacco Research*, 5(4), pp. 575-584.
- Hodges, J.Q. (2016), "Predicting Use of Community Mental Health Services: Do Demographics of Consumers Matter?", *Journal of Sociology and Social Work*, 4(1), pp. 42-49.
- Hughes, A.O., Fenton, S., and Hine, C.E. (1995), "Strategies for Sampling Black and Ethnic Minority Populations", *Journal of Public Health Medicine*, 17(2), pp. 187-192.
- Kalton, G., and Anderson, D. W. (1986), "Sampling Rare Populations", *Journal of the Royal Statistical Society, Series A (General)*, 149(1), pp. 65-82.
- Leonard, M.M. (2021), "Sampling Portuguese Migrants and Their Descendants in Germany", paper presented at the German Political Science Association (DVPW), online.
- Lindsay, A.C., Wallington, S.F., Rabello, L.M., et al. (2020), "Faith, Families, and Social Networks: Effective Strategies for Recruiting Brazilian Immigrants in Maternal and Child Health Research", *Journal of Racial and Ethnic Health Disparities*, DOI:10.1007/s40615-020-00753-3.
- Lohr, S.L. (2021), "Multiple-Frame Surveys for a Multiple-Data-Source World", *Survey Methodology*, 47(2), pp. 229-263.
- Lynn, P., Nandi, A., Parutis, V., and Platt, L. (2018), "Design and Implementation of a High-Quality Probability Sample of Immigrants and Ethnic Minorities: Lessons Learnt", *Demographic Research*, 38(21), pp. 513-548.
- Marpsat, M., and Razafindratsima, N. (2010), "Survey Methods for Hard-to-Reach Populations: Introduction to the Special Issue", *Methodological Innovations Online*, 5(2), pp. 3-16.
- McDougall, G.J., Holston, E.C., and Wilke, P. (2001), "Recruiting African Americans into Research on Cognitive Aging", *Ethnicity & Disease*, 11, pp. 124-133.
- McKenzie, D.J., and Mistiaen, J. (2007), "Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys", IZA Discussion Paper. No. 3173, Bonn, Germany: Institute for the Study of Labor.
- Nam, Y., Mason, L. R., Kim, Y., Clancy, M., and Sherraden, M. (2013). "Survey Response in a Statewide Social Experiment: Differences in Being Located and Collaborating, by Race and Hispanic Origin", *Social Work Research*, 37(1), pp. 64-74.
- Rao, J.N.K. (2021), "On Making Valid Inferences by Combining Data from Surveys and Other Sources", *Sankhyā, Series B*, 83-B, pp. 242-272.
- Reichel, D., and Morales, L. (2017), "Surveying Immigrants without Sampling Frames – Evaluating the Success of Alternative Field Methods," *Comparative Migration Studies*, 5(1), DOI:10.1186/s40878-016-0044-9

- Simon, P. (2012), "Collecting Ethnic Statistics in Europe, A Review", *Ethnic and Racial Studies*, 35(8), pp. 1366-1391.
- Statistisches Bundesamt [Destatis] (2022), "Population by Migrant Status and Sex", Retrieved: <https://www.destatis.de/EN/Themes/Society-Environment/Population/Migration-Integration/Tables/migrant-status-sex.html#268892>
- Sudman, S., and Kalton, G. (1986), "New Developments in the Sampling of Special Populations", *Annual Review of Sociology*, 12, pp. 401-429.
- Tourangeau, R., Edwards, B., Johnson, T., Wolter, K., and Bates, N. (eds) (2014), *Hard-to-Survey Populations*, Cambridge: Cambridge University Press.
- Vu, M., Huynh, V.N., Bednarczyk, R.A., et al. (2021), "Experience and Lessons Learned from Multi-Modal Internet-Based Recruitment of U.S. Vietnamese into Research", *PLoS ONE*, 16(8):e0256074, DOI:10.1371/journal.pone.0256074.
- Weinmann, T., AlZahmi, A., Schneck, A., et al. (2019), "Population-Based Assessment of Health, Healthcare Utilisation, and Specific Needs of Syrian Migrants in Germany: What is the Best Sampling Method?", *BMC Medical Research Methodology*, 19(5), DOI:10.1186/s12874-018-0652-1.
- Willis, G.B., Smith, T.W., Shariff-Marco, S., and English, N. (2014), "Overview of the Special Issue on Surveying the Hard-to-Reach", *Journal of Official Statistics*, 30(2), pp. 171-176.
- Zemore S.E., Gilbert, P.A., Pinedo, M., et. al. (2021), "Racial/Ethnic Disparities in Mutual Help Group Participation for Substance Use Problems", *Alcohol Research*, 41(1), DOI:10.35946/arcr.v41.1.03.