SURVEYS · SAMPLING

Handling Complex Household Structures in Survey Sampling Frames

Most general population surveys sample addresses but want to measure people. This article explains how dwellings, households and persons are linked, where complex living arrangements cause coverage and selection errors, and how to handle them.

ModQu research article · 2,355 words · 22 references (APA 7th edition)

Most surveys of the general population do not sample people directly. In many countries no complete, accessible list of residents exists, so surveys sample addresses or dwellings and then identify the people who live there. The chain from address to person seems simple: select an address, find the household, list its members and interview one or all of them. In practice each link can break. One address may contain several dwellings, one dwelling several households, and some people belong to more than one household or to none. These complexities create coverage errors and unequal selection probabilities that, if ignored, bias survey estimates.

This article explains the units involved, the problems complex household structures cause and the methods that survey designers use to deal with them.

Addresses, dwellings, households and persons

Four different units are involved, and confusion between them is a frequent source of error (Kish, 1965).

- An **address** is a location on a postal or administrative list.
- A **dwelling** or **housing unit** is a separate set of living quarters, usually with its own entrance and cooking facilities.
- A **household** is a group of people who share a dwelling and, under many definitions, share meals or a common budget.
- A **person** is the unit about whom most survey estimates are made.

The **sampling frame** usually lists addresses. Address-based sampling, using postal delivery files, has become the main approach for general population surveys in several countries because it covers most residential addresses and allows surveys to be conducted by mail, web or in person (Iannacchione, 2011; Link et al., 2008). Area frames, in which interviewers list the dwellings in selected small areas, remain important where no good address list exists.

Figure 1 shows the chain from frame to respondent and the problems that can occur at each link.

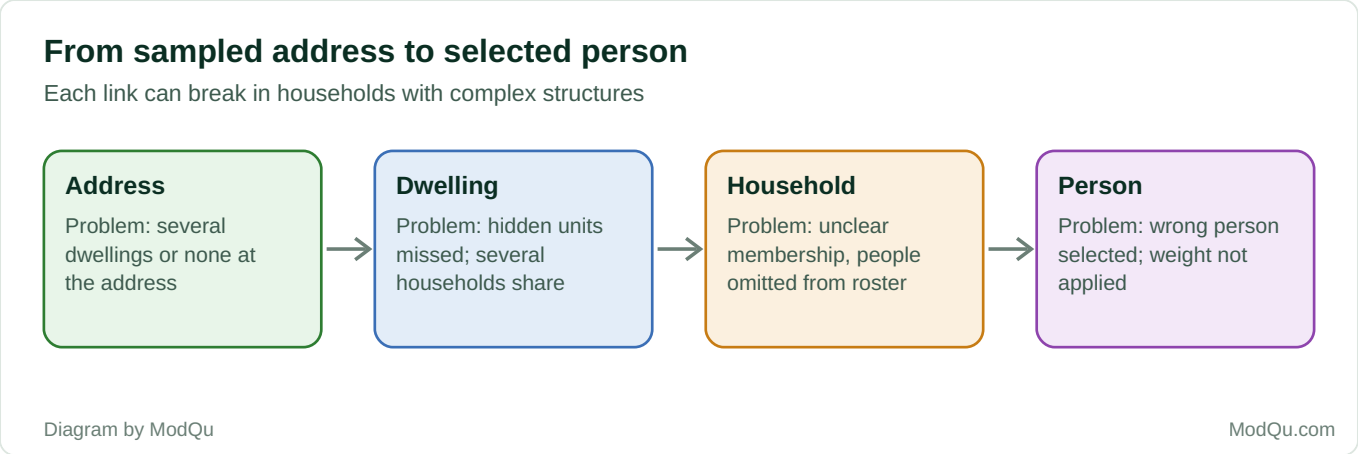


Figure 1. From sampled address to selected person, and the problems at each link. Diagram by ModQu.

Problems at each link

Address to dwelling

An address may contain more than one dwelling. Basements, converted houses, rooms above shops and informal secondary units may have no separate address. If interviewers or respondents treat the address as a single dwelling, the extra units are missed. Conversely, some addresses contain no dwelling at all, because they are businesses, vacant or demolished.

Area listing has its own problems. Eckman and Kreuter (2011) found that interviewers who were given an existing list to update tended to confirm the units on the list rather than add missing ones, a form of confirmation bias. Eckman (2013) found that different listers working in the same areas produced noticeably different lists. Hidden and irregular dwellings are the ones most often missed, and their residents tend to differ from the rest of the population.

Practical safeguards include instructions to ask at each address whether there are other places where people live, such as a basement flat or a separate unit behind the house, and procedures for adding such units to the sample with a known probability.

Dwelling to household

Most surveys assume one household per dwelling, but shared dwellings are common among young adults, migrants, students and people on low incomes. Whether these residents form one household or several depends on the definition used. The United Nations (2005) describes the two main concepts: the *housekeeping unit*, defined by shared meals or budgets, and the *household-dwelling* concept, under which everyone in a dwelling forms one household. The choice affects household counts, household income and the selection of respondents.

The survey's definition must be written into the questionnaire and applied consistently. If several households share a dwelling, the survey must either include all of them or select among them with known probabilities.

Household to person: who lives here?

Listing the members of a household, the **household roster**, is the step where complex structures cause the most trouble. Respondents do not always know whom to include, and survey rules about "usual residence" may conflict with their own sense of who belongs.

Tourangeau et al. (1997) found that undercoverage in household surveys arose partly because respondents left people off rosters, either deliberately, to conceal them, or because the rules were unclear. Martin (2007) examined people with tenuous ties to residences, such as those who move between several homes, stay temporarily with friends or relatives or are away for work. She found that such people were more likely to be omitted, and that coverage improved when roster questions asked explicitly about such cases.

Common difficult cases include:

- **Children in shared custody** who spend part of the week in each parent's home.
- **Students** who live away during term but return home in holidays.
- **Commuter workers** who keep a second residence near their job.
- **People staying temporarily**, including those with no other home.
- **People away in institutions**, such as hospitals, care homes, prisons or the armed forces.
- **Lodgers, boarders and domestic workers** who live in but may not be considered family.
- **Multigenerational and blended households**, in which several family units share a home.

Clear residence rules, examples on the questionnaire or interviewer screen, and probing questions about people who are sometimes there reduce omissions. Each person should belong to exactly one household under the rules, so that nobody is counted twice or missed.

A worked example

Consider a hypothetical house selected from an address list. The person who answers the door reports that she lives there with her husband and their two children. Probing reveals more. Her mother has lived with them since last year. A lodger rents a room and buys his own food. Their son is at university in another city during term. A daughter from her husband's earlier marriage stays every second weekend. A friend has been sleeping on the sofa for three weeks while looking for a flat. And a separate flat in the basement, with its own door at the side of the house, is rented to a young couple.

Without clear rules, different interviewers might record anything from four to eleven residents in one or two households. Under typical survey rules, the basement flat is a separate dwelling and should be added to the sample as such. The mother is a usual resident. The lodger may be a separate household under a shared-meals definition but part of the household under a dwelling definition. The student is usually counted at the term-time address. The stepdaughter is counted where she spends most nights. The friend's status depends on whether he has another usual residence. Each decision changes the household's size, composition and income, and the probability that each adult is selected.

People outside households

Some people do not live in private households at all. Residents of care homes, student halls, prisons, military barracks and shelters, and people without housing, are usually excluded from household frames. Surveys must state these exclusions, and estimates should not be described as covering "the population" if they cover only the household population. Where these groups matter, separate frames and methods are needed (Kalton & Anderson, 1986).

Within-household selection

When a survey interviews one adult per household, it must choose which adult. Letting the person who answers the door or the telephone respond is quick but biased, because the people most often at home or most willing to answer differ from others. Survey methodologists have developed several selection methods.

The Kish grid. Kish (1949) proposed listing all eligible adults in a set order, usually by sex and age, and selecting one using a table pre-assigned to each address. The method gives every adult a known probability of selection and is close to a true random selection, but it requires a full roster, which takes time and can feel intrusive.

Birthday methods. The respondent is the adult with the next or most recent birthday. These methods are quick and less intrusive, but they rely on the informant knowing every adult's birthday and following the instruction.

Age and sex quota methods. Methods such as asking for the youngest man or, if none, the oldest woman are quick but do not give known probabilities.

Gaziano (2005) compared many selection methods and concluded that none was best in every respect: the more rigorous methods produced better samples but at a higher cost in time and cooperation. In mail surveys, where no interviewer is present, compliance is a particular problem. Battaglia et al. (2008) and Olson et al. (2014) found that a substantial share of household mail survey respondents were not the person who should have been selected under the birthday method. Errors were more common in larger households and in households with particular compositions, which means that selection errors are not random.

Self-administered modes

Rosters and selection are harder without an interviewer. In web and mail surveys, the person who opens the invitation must read the residence rules, list or count the household's adults and pass the questionnaire to the right person. Each step depends on the respondent's understanding and willingness. Long rosters increase burden and break-off, so self-administered surveys often ask only for the number of adults and use a birthday rule. Some designs avoid selection entirely by inviting every adult in the household to respond, accepting the clustering of answers within households in return for simpler instructions.

Weighting for household size

If one adult is selected from each sampled household, adults in larger households have a lower chance of selection. An adult living alone is certain to be selected if the household is sampled; an adult in a four-adult household has a one-in-four chance. To produce unbiased estimates for adults, each respondent must be weighted by the number of eligible adults in the household, the inverse of their within-household selection probability (Kish, 1965; Lohr, 2022).

$$w_i = w_h \times a_h$$

Here w_h is the base weight of the sampled household and a_h is the number of eligible adults in that household.

Figure 2 shows what happens if this weight is ignored, using a hypothetical population. Adults in one-adult households make up 30 per cent of the unweighted sample but only about 15 per cent of the adult population. Any characteristic linked to living alone, such as age, income or social contact, would be misestimated.

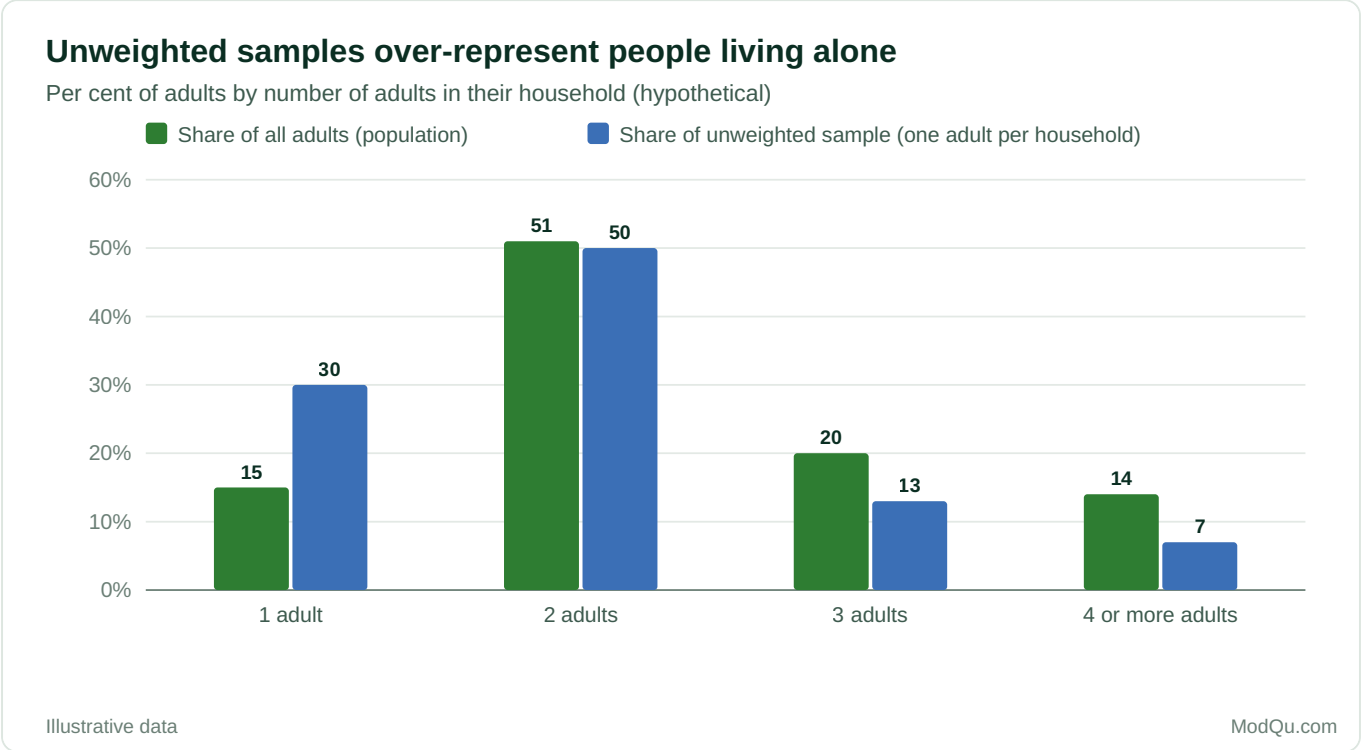


Figure 2. Why within-household selection needs weighting. Hypothetical population in which 30, 50, 13 and 7 per cent of households have one, two, three and four adults. Illustrative data; chart by ModQu.

The weight increases the variation in weights and so reduces precision. Designers sometimes cap the adjustment or interview more than one adult in large households to limit this loss, while keeping probabilities known (Valliant et al., 2018).

Household-level estimates need different weights. If the survey also produces estimates about households, such as the share of households with a given appliance or the average household income, the household base weight is used without the adjustment for the number of adults. Analysts must therefore use the correct weight for each type of estimate.

Multiplicity and linked people

Some research questions concern people linked to households in more than one way. Sirken (1970) introduced **multiplicity sampling**, in which sampled households report not only on their own members but also on relatives living elsewhere, such as siblings or children. A person can then be reached through several households, and their selection probability depends on how many households could report them. Estimation weights each person by the inverse of this multiplicity. The approach is useful for rare events and populations but depends on accurate reporting of the links.

More generally, Lavallée (2007) developed **indirect sampling**, in which the population sampled (such as dwellings) differs from the target population (such as people or families), and links between the two are used to calculate weights. Indirect sampling gives a formal framework for handling complex structures, including people who belong to more than one household.

Frames that list people

In some countries, survey designers can sample people directly from population registers or electoral rolls. Person frames avoid within-household selection but bring their own problems: registers may be out of date, may omit recent movers and may list people at addresses where they no longer live. Lynn and Taylor (1995) compared samples of individuals drawn from British electoral registers with samples drawn from the Postcode Address File. They found differences in coverage, with the address file performing better for groups less likely to be registered, and differences in the variance of estimates because of clustering and weighting. The choice of frame thus affects both bias and precision.

Implications for survey error

Problems with complex households affect several components of total survey error. Missing dwellings and omitted household members cause coverage error. Incorrect within-household selection behaves like nonresponse or selection bias. Errors in counting adults distort weights. Groves (1989) emphasised that coverage error is often hard to detect because the missing people leave no trace in the data. Groves and Couper (1998) showed that household characteristics, such as size and composition, are also linked to contact and cooperation, so complex households may be underrepresented through nonresponse as well as coverage.

Practical guidance

1. **Define the units.** Write clear definitions of dwelling, household and eligible person, and apply them consistently.
2. **Check for hidden units.** Ask at each address whether other people live in separate quarters.
3. **Write residence rules for difficult cases.** Cover shared custody, students, commuters, temporary residents and people away in institutions, with examples.
4. **Probe the roster.** Ask whether anyone else usually lives there or stays there sometimes.

5. **Use a selection method with known probabilities.** Choose the most rigorous method the mode and budget allow, and monitor compliance.
6. **Weight for selection.** Include the number of eligible adults in the weight, and record it accurately.
7. **State exclusions.** Report which populations are outside the frame.
8. **Test.** Pretest roster questions with people living in complex arrangements (Groves et al., 2009; Kalton, 1983).

Key takeaways

Household surveys measure people through a chain of units: addresses, dwellings, households and persons. Complex living arrangements weaken each link. Hidden dwellings, shared homes, people with tenuous ties to residences and unclear residence rules cause coverage errors, while within-household selection and weighting determine whether the people interviewed fairly represent all adults. Careful definitions, probing roster questions, rigorous selection and correct weights turn an address sample into a sound sample of people.

References

Battaglia, M. P., Link, M. W., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2008). An evaluation of respondent selection methods for household mail surveys. *Public Opinion Quarterly, 72*(3), 459–469. https://doi.org/10.1093/poq/nfn026

Eckman, S., & Kreuter, F. (2011). Confirmation bias in housing unit listing. *Public Opinion Quarterly, 75*(1), 139–150. https://doi.org/10.1093/poq/nfq066

Eckman, S. (2013). Do different listers make the same housing unit frame? Variability in housing unit listing. *Journal of Official Statistics, 29*(2), 249–259. https://doi.org/10.2478/jos-2013-0021

Gaziano, C. (2005). Comparative analysis of within-household respondent selection techniques. *Public Opinion Quarterly, 69*(1), 124–157. https://doi.org/10.1093/poq/nfi006

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. Wiley. https://doi.org/10.1002/9781118490082

Groves, R. M. (1989). *Survey errors and survey costs*. Wiley. https://doi.org/10.1002/0471725277

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Iannacchione, V. G. (2011). The changing role of address-based sampling in survey research. *Public Opinion Quarterly, 75*(3), 556–575. https://doi.org/10.1093/poq/nfr017

Kalton, G., & Anderson, D. W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society: Series A (General), 149*(1), 65–82. https://doi.org/10.2307/2981886

Kalton, G. (1983). *Introduction to survey sampling*. SAGE. https://doi.org/10.4135/9781412984683

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association, 44*(247), 380–387. https://doi.org/10.1080/01621459.1949.10483314

Kish, L. (1965). *Survey sampling*. Wiley.

Lavallée, P. (2007). *Indirect sampling*. Springer. https://doi.org/10.1007/978-0-387-70782-2

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2008). A comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly, 72*(1), 6–27. https://doi.org/10.1093/poq/nfn003

Lohr, S. L. (2022). *Sampling: Design and analysis* (3rd ed.). CRC Press. https://doi.org/10.1201/9780429298899

Lynn, P., & Taylor, B. (1995). On the bias and variance of samples of individuals: A comparison of the Electoral Registers and Postcode Address File as sampling frames. *Journal of the Royal Statistical Society: Series D (The Statistician), 44*(2), 173–194. https://doi.org/10.2307/2348443

Martin, E. (2007). Strength of attachment: Survey coverage of people with tenuous ties to residences. *Demography, 44*(2), 427–440. https://doi.org/10.1353/dem.2007.0011

Olson, K., Stange, M., & Smyth, J. (2014). Assessing within-household selection methods in household mail surveys. *Public Opinion Quarterly, 78*(3), 656–678. https://doi.org/10.1093/poq/nfu022

Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the American Statistical Association, 65*(329), 257–266. https://doi.org/10.1080/01621459.1970.10481077

Tourangeau, R., Shapiro, G., Kearney, A., & Ernst, L. (1997). Who lives here? Survey undercoverage and household roster questions. *Journal of Official Statistics, 13*(1), 1–18.

United Nations. (2005). *Household sample surveys in developing and transition countries* (Studies in Methods, Series F, No. 96). United Nations.

Valliant, R., Dever, J. A., & Kreuter, F. (2018). *Practical tools for designing and weighting survey samples* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-93632-1