How to design and implement surveys that are fit-for-purpose

How to design and implement surveys that are fit-for-purpose

Peter Lynn

Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Nature Human Behaviour* following peer review. The version of record (published version) [How to design and implement surveys that are fit-for-purpose*;* Peter Lynn; Nature Human Behaviour 2025; doi: 10.1038/s41562-025-02317-z] will be available online at: *https://www.nature.com/nathumbehav/*

and should be cited as follows:

Lynn, Peter (2025) 'How to design and implement surveys that are fit-for-purpose'. *Nature Human Behaviour,* published online 1 October 2025.

Author contact details: Institute for Social and Economic Research, University of Essex,

Wivenhoe Park, Colchester, Essex CO4 3SQ; e-mail: plynn@essex.ac.uk.

How to design and implement surveys that are fit-for-purpose

Peter Lynn

Survey methods are used to collect informative data in a wide range of scientific contexts. This article outlines how to ensure that a survey is fit for its intended purpose and to avoid the many potential pitfalls associated with survey research.

Survey or questionnaire-based research can be a powerful tool to provide quantitative information about a study population. Survey methods can obtain information that cannot be obtained by other means, such as opinions, values, expectations, and the reasons for making decisions, and information pertaining to a particular population of interest, about which little may be known. If well-designed (see Box 1), a survey can provide estimates of the characteristics of the population of interest with known precision, thereby enabling hypothesis testing and model fitting.

However, care must be taken to ensure that the survey provides data that are sufficiently relevant, precise, unbiased and reliable to support the analysis purposes for which it is intended. There are many possible pitfalls that can cause survey data to be misleading. But, with a little care and attention to the way a survey is designed and implemented, these pitfalls can usually be avoided. This article explains how.

The Sample

To provide a meaningful basis for inference and to avoid bias, a survey sample must represent the target population of inference. This means it must share all relevant characteristics of that population. A sample must additionally be designed to provide sufficient precision (low sampling variance) to meet the survey aims. Sample size is a key driver of precision, but not the only one.

The choice of sampling method is important as this can affect the sample composition and the likelihood of bias. Some commonly used methods are visualised in Figure 1 and explained in what follows. Volunteer samples and convenience samples are best avoided. It is unlikely that volunteers will have the same attitudes and experiences as others. People tend to volunteer if they have strong opinions on, or are particularly interested in, the topic of the survey.

A commonly used non-representative sampling method is quota sampling. Quota sampling ensures the sample reflects the population distribution of a small number of variables such as age, sex and employment status. While arguably better than an uncontrolled volunteer sample, this is still essentially a volunteer sample. The sample will remain skewed towards the most willing and available respondents and - unless the

quota variables explain almost all the variance in key outcomes of interest, which is unlikely – may therefore be biased in terms of key variables and outcomes[1].

It is preferable to control sample selection in an objective way that seeks to avoid bias. In most contexts, representative sampling is perfectly possible though it may be a little more time consuming or more expensive than accepting a quota sample or other form of volunteer sample. The essence of objective sample selection is to not allow any person – whether that is the potential sample member or a field worker recruiting the sample – any possibility to introduce subjectivity into the process, but instead to have selection governed by a random chance mechanism.

For example, if the aim is to collect data from hospital patients attending an outpatient appointment, a random sample could consist of every n^{th} patient attending. This requires the sample identification process either to be programmed into the registration system or administered manually by reception staff, which is of course somewhat more demanding and costly than simply posting an open invitation on the wall of the waiting room asking for volunteer respondents, but the extra effort may be easily justified from a value-for-money perspective. For surveys of the general population, the best approach is to select a random sample of persons from a population register or other register with good coverage, if this is available, or alternatively to randomly select residential addresses. In the latter case, a subsequent stage of selection of one or more eligible persons at the address is also needed. This too should be done using a controlled random mechanism, which can be achieved whether the survey is being administered by interviewers or is self-administered[2].

The most basic form of random sampling is 'simple random sampling', in which each population unit has an equal and independent chance of being selected. An improvement on this is 'stratified random sampling', which ensures the sample reflects the population distribution across strata while maintaining the probability basis for selection. The stratification can be either implicit or explicit. 'Implicitly stratified random sampling' consists of sorting a list of population units in a meaningful way and then selecting from the ordered list using a random start and a fixed interval. Selecting every n^{th} attendee at an outpatient clinic is an example, where the sorting principle is time and day of attendance. If appointment timing has systematic tendencies, this improves statistical precision. 'Explicitly stratified random sampling' consists of sorting population units into explicit groups (strata) and selecting a random sample independently from each stratum. Either form of stratification can often be introduced with relatively little cost and effort but with useful gains to statistical precision[3].

Sometimes, 'multi-stage sampling' is an effective way to control the costs and difficulties of data collection. The first-stage sampling units (clusters) may be hospitals, schools, or small geographical areas, for example, depending on the context. However, it remains highly advantageous to use probability-based selection methods at each stage. Multi-

stage sampling tends to reduce the precision of sample-based estimates, other things being equal. This is because statistically it has the opposite effect of stratification: instead of ensuring the sample is spread evenly over all population subgroups (strata), it causes the sample to be restricted to a limited number of subgroups (clusters). This precision reduction is usually easily justified in value-for-money terms.

The Questionnaire

Another central feature of survey research is the questionnaire. An effective questionnaire will strike a balance between the desire of the researcher to obtain detailed information on many related issues and the ability and willingness of the respondent to provide accurate information. A survey should have clear objectives in terms of concepts to be captured to support the analysis plan, and the questionnaire content should reflect this.

A well-designed questionnaire will communicate clearly to the respondent what is required and will ask only for information that the respondent can reasonably be expected to know, in terms that they will understand. The questionnaire should use clear, consistent wording, should be logical and intuitive in structure, and should encourage honest and accurate answers[4,5].

A common temptation is to expand a questionnaire beyond that which is needed to fulfil the core objectives of a study. While this may in principle open up additional avenues of research and provide additional insights, it also risks burdening the respondent unduly, which could result either in hasty, imprecise answers or, worse, in the respondent deciding not to complete the questionnaire.

A survey questionnaire should always be tested before deployment. Testing methods need not be unduly expensive or time consuming. Simple methods such as expert review and cognitive testing (get three or four people to fill the questionnaire and provide feedback on the experience) can reduce the risk of poor-quality data[6].

There are different ways to administer questionnaires. Self-completion methods (online or on paper) generally elicit more honest answers[7] and are less costly, while interviewer-administered methods can improve co-operation rates and can support the collection of more complex information and the use of longer lists of response options. If the questionnaire needs to include a request for consent to some kind of follow-up intervention, or to link administrative data to the survey responses, consent rates tend to be much higher with interviewer administration[8].

Obtaining Responses

A well-designed questionnaire is not much use unless study respondents complete it. Small things can make a difference to the quantity and quality of responses received, such as the day of the week on which the survey invitation is sent[9], the number and nature of reminders that are sent (vary them: a letter, a postcard, a paper version of the questionnaire)[10], and the persuasive arguments put forward in the invitation letter[11]. Small incentives such as cash or a cash-equivalent gift will improve response rates[12].

The various design tools available to survey researchers to help maximise the chances that sample members participate tend to have heterogeneous effects across population subgroups. Incentives will be effective for some, while appeals to altruism will work for others, for example. Consequently, survey designers have begun to use targeted designs, in which different combinations of relevant design features are administered to different sample subgroups, based on prior knowledge of likely effectiveness[13]. Even basic information about sex and age, or region of residence, can be enough to enable effective targeting strategies[11,13], which can be no more costly than standardised approaches. Knowledge about effective targeted design features for different situations can be found at https://surveyfutures.net/wp-content/uploads/2025/05/report-5-targeted-procedures-evidence-review.pdf.

In the case of longitudinal studies, in which questionnaires are administered repeatedly to sample members, dropout can be reduced by engendering a sense of study loyalty through branding, communications, and feedback. Consistent messaging and terminology should be used at each survey wave and additional contacts between waves may be desirable if the interval between waves is several months or more. These occasions can be used to keep in touch and to motivate continued participation by feeding back results from the study.

Analysis and Estimation

Unless a simple random sample is used, analysis of survey data should take into account the sample design. If selection probabilities differed between strata or subgroups, design weights should be applied to correct the imbalance[14]. For example, if the proportion of older people sampled was twice as greater as that of younger people, each older respondent should be given half the weight of each younger respondent in analysis if results are to represent the population. An adjustment to these weights to correct for differential effects of nonresponse is typically also advisable[15]. For example, if the response rate was 50% amongst men and 67% amongst women, design weights should be multiplied by 2.0 for men and by 1.5 for women. For longitudinal surveys, the adjustment may need to be made after each wave of data collection, if analysis is to proceed after each wave, or just once after the survey is completed if analysis is to

proceed only once all data is collected[16]. If the sample design is multi-stage or involves stratification, appropriate variance estimation methods should be used[17]. Without this, standard errors and hypothesis tests will be biased – possibly severely so. Many standard statistical software packages – such as Stata, SPSS and SAS - incorporate appropriate routines for survey variance estimation, while user-produced syntax is available in R.

Conclusion

Survey research can be highly valuable, but only if conducted to an appropriate standard that produces data fit to meet the study objectives. High quality survey research does not necessarily require large budgets and considerable resources. It requires careful attention to each stage of the survey process, namely how the sample is designed and collected, how the questionnaire is designed, how respondents are encouraged to participate, and how the data are analysed. Lack of attention at any one of these stages can undermine the utility of the survey.

Box 1

How to Conduct Good Surveys

The sample should be representative of the population to which you want to make inferences. The best way to ensure this is to make selections using an objective, probability-based, method.

Questionnaire design should follow established principles of clarity, specificity, simplicity and applicability. Questionnaires should be tested prior to survey implementation and should appear attractive, relevant and non-burdensome.

Sample members should be encouraged to respond through appeals to both altruism and egoism. Participation should be convenient and cost-free to the respondent and reminders should be deployed.

Estimation should be based on methods that take into account the sample design and correct for the effects of non-response.

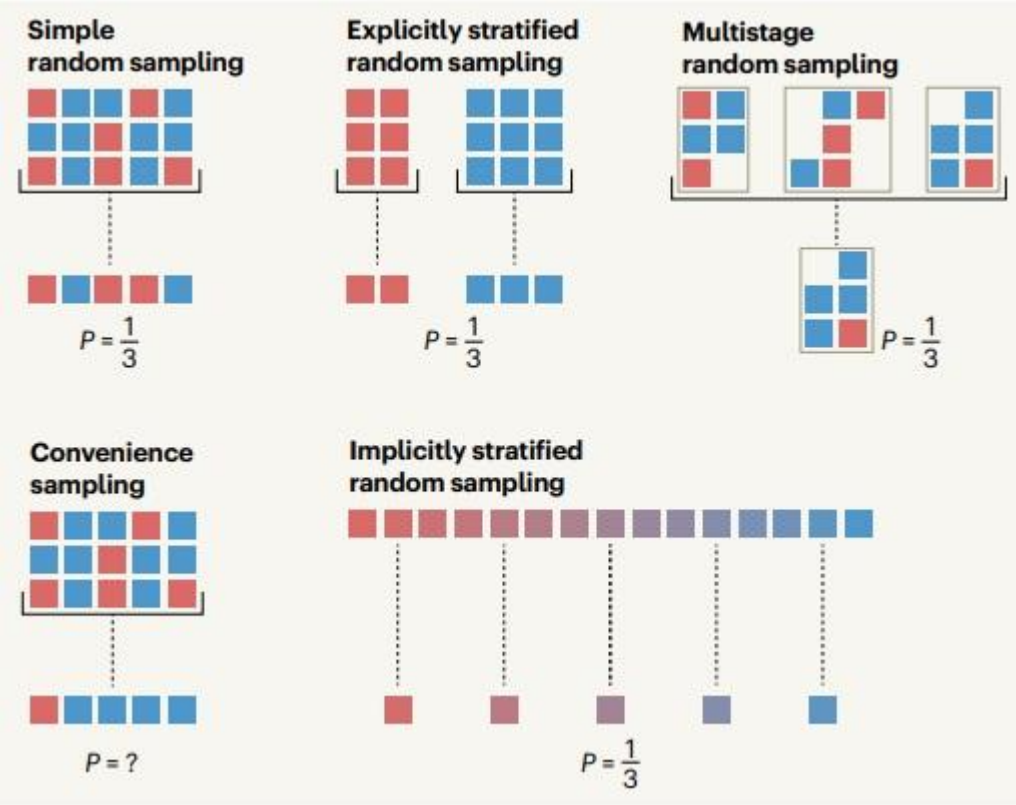

Figure 1. Overview of sampling methods and samples

Competing interests

The author does not declare any competing interests.

References

1. Brick, M. *Pub Op Qtly.* **75,** 872–888 (2011).
2. Smyth, J.D., Olson, K. & Stange, M. in *Experimental Methods in Survey Research* (ed.s Lavrakas, P.J. et al.) 23-45 (Wiley, 2017).
3. Lynn, P. *Methods, Data, Analyses.* **13**, 253-266 (2019).
4. Fowler, F. J. Jr. *Improving Survey Questions: Design and Evaluation* (Sage, 1995).
5. Tourangeau, R., & Yan, T. *Psych Bull*. **133,** 859-883 (2007).
6. Presser, S. et al. (Eds.) *Methods for Testing and Evaluating Survey Questionnaires*. (Wiley, 2004).
7. Tourangeau, R., & Yan, T. *Psych Bull*. **133,** 859-883 (2007).
8. Jäckle, A., Burton, J., Couper, M.P., Crossley, T.F. & Walzenbach, S. *Surv Res Meth*. **16,** 387-408 (2022).
9. Lynn, P., Bianchi, A. & Gaia, A. *Soc Sci Comp Rev*. **42,** 352-368 (2024).
10. Cabrera Alvarez, P. & Lynn, P. *Jnl Surv Stat Meth*. **12,** 651-673 (2024).
11. Lynn, P. *Pub Op Qtly*. **80,** 771-782 (2016).
12. Singer, E. & Ye, C. *Ann Am Acad Pol Soc Sci*, **645**, 112-141 (2012).
13. Lynn, P. *Surv Res Meth*, **11**, 93-103 (2017).
14. Lynn, P. in *Encyclopedia of Social Measurement* (ed. Kempf Leonard, K.) 967-973 (Academic Press, 2005).
15. Lynn, P. in *Survey and Statistical Computing 1996* (eds Banks, R. et al) 205-214 (Association for Survey Computing, 1996).
16. Lynn, P. & Watson, N. in *Advances in Longitudinal Survey Methodology* (ed. Lynn, P.) 447-468 (Wiley, 2021).
17. Valliant, R., Dever, J.A. & Kreuter, F. *Practical Tools for Designing and Weighting Survey Samples*, 2nd edn (Springer, 2018).