ANNEX: FURTHER COMMENTS ON SURVEYS

SAMPLING METHODS

Broadly, sampling methods are typically put into one of two categories: probability or non-probability sampling.

Probability sampling involves selecting your sample using random or probabilistic methods.

A common probability sampling method is simple random sampling, which involves randomly selecting individuals from the sample frame such that everyone has the same chance of being surveyed. Implementing simple random sampling is straightforward.

An alternative is stratified random sampling, which involves first splitting the sample frame into groups (strata) based on specified characteristics like ethnicity or membership type, then randomly sampling from each stratum such that the final sample is proportional to the population according to the specified characteristics. It is common to stratify based on demographic characteristics, but you may consider stratifying by membership class or duration of membership, among other options.

Non-probability sampling methods, on the other hand, do not select a sample based on randomised or probabilistic procedures. Instead, these methods prioritise getting responses from members that are the most convenient to reach, or make use of social networks to find and recruit interested members for participation.

Generally, if you are interested in understanding the views of your whole membership or large groups within it, then you should avoid using non-probability sampling techniques, because they are unlikely to give you a representative perspective.

However, you might consider using non-probability sampling methods where it is otherwise not possible to access certain kinds of members. This could be the case if you are interested in conducting extended interviews with members from a very small minority group for which you do not collect identifying information elsewhere via a CRM. In those cases, it might be helpful to use what's called "snowball sampling," which involves taking advantage of the social links between interviewees to boost the overall number of people you can contact.

SAMPLE SIZES AND RESPONSE RATES

The first thing many organisations think about when it comes to surveys is the sample sizes and response rates that they generate. There is often a lot of concern given to whether these are "good" or "bad," but it is helpful to clarify why they are important.

The purpose of asking a question on a survey is to get an understanding of how much of your population has a certain characteristic, what proportion holds a certain opinion, etc. You estimate these parameters of your population using the responses to your survey. That is, if 25% of your respondents said that they dislike your current renewal process, then you estimate that 25% of your population hold the same view. However, there are two main points you should think about when considering the estimates that you derive from your surveys.

Margin of Error

The first of these has to do with the precision of your estimates. You can easily imagine that a different sample responding to the same survey would produce a different estimate of the population parameter. Thus, every estimate derived from a survey sample is inherently uncertain. We quantify that uncertainty by calculating a margin of error, which is usually expressed like "+/- 3%," indicating that we expect the population parameter to be within three percentage points of our estimate.

For example, a survey might show that 60% of members who responded to a survey felt satisfied with their membership. Assuming that the sample is broadly representative of the population along relevant characteristics

and our survey is free from various sources of error[1], we can reliably estimate that 60% of the full membership are also satisfied. However, we also calculate a margin of error around that estimate. If the margin of error is 2.5%, then we conclude that, if we were to repeat the survey many times, then there is a 95% probability that the true satisfaction rate for the full membership is between 57.5% and 62.5%.[2]

The margin of error will become smaller as sample sizes increase. This means we can make more precise inferences about the likely true value we are trying to estimate for the population. It is often more helpful to think about the margin of error than the sample size itself, because for practical purposes what is important is whether we can reasonably rule out certain conclusions.

Consider a survey in which you ask: "Do you agree or disagree that Equity, Diversity, and Inclusion (EDI) should be prioritised by the organisation?" When you start to analyse the data, you find that members agree 50% of the time, disagree 40% of the time, and are unsure or prefer not to say 10% of the time. You calculate that for this sample you have a relatively high margin of error of +/- 10%.

Here, you might be hesitant to make strong moves towards a new EDI initiative, not only because the difference between agreeing and disagreeing groups is relatively small, but also because your estimates are rather imprecise—it would not be unreasonable to ask whether the true views among your members are closer to 40% in agreement and 50% in disagreement, or even 60% in agreement and 30% in disagreement. However, if you collected more data, you might become more confident about whether your members believe EDI should be a priority. A "good" sample size is one that gives you enough precision for your estimates to be of practical use.

Conventionally, a margin of error of 2.5 percent is considered "good" by public pollsters, but larger margins of error can be acceptable in other contexts.

Sources of Bias

The second point you need to think about is the various sources of systematic error that can affect your estimates. In essence, this has to do with the *accuracy* of your estimates. There are many possible sources of error, and a complete discussion would be lengthier and more technical than would be appropriate for this report. Instead, consider just one possible source of error that has particular relevance for response rates: non-response bias. This is a form of error that arises when survey non-respondents would systematically give different answers than respondents. If there is non-response bias, then your estimates will be biased in a way that can make them inaccurate representations of the population.

For the current discussion, the important thing to note is that non-response bias is not necessarily higher when response rates are lower. Rather, the *possible magnitude* of the non-response bias is higher the lower the response rate. In other words, we cannot necessarily say that a higher response rate implies a lower non-response bias. Rather, we can just be more confident that its magnitude is lower than it otherwise could be.

Therefore, although response rates are important, you should reinvest some thinking about response rates into reasoning about ways in which your results could be biased. A great deal of this work should be done in the survey planning and design stages so that you minimise error before it can crop up in your survey. Still more work can be done during the analysis stage if you are able to weight responses to achieve proportionality in characteristics you expect to be important to the results. For instance, you might apply weights based on membership class when estimating the overall opinion of your members about some topic, based on previous research that found substantial differences between each of your member classes on a related question. You

---

[1] These assumptions will be more or less reasonable in different circumstances depending on the sampling method, survey question, the population being examined, etc. We discuss these in the next subsection, but in general we encourage organisations to seek out statistical expertise when designing, implementing, and interpreting the results of surveys.

[2] This assumes a 95% confidence level used in the calculation of the margin of error. A higher or lower level could be used, which would result in a larger or smaller margin of error, respectively.

should also try to communicate the limitations of your surveys when reporting the results. This should include consideration of where bias might affect things and the steps you were able to take towards mitigating bias.

Where do diversity characteristics come into the picture? If there is a correlation between diversity characteristics and responses to a survey question, then having respondents that are not representative along those characteristics risks biasing your results for that question. If you know the characteristics of respondents and non-respondents, you can compare them to your population to see if there are correlations between their characteristics and the rate at which they respond or do not respond.

Having a set of respondents that are broadly reflective of your population might be a sign that your estimates are reliable, but it is no guarantee. It may be that non-responding members still hold different views than respondents, even if their known characteristics align proportionately. In general, there are no ways to fully ensure the absence of non-response bias. In practice, it is common to use things like the demographic characteristics of your respondents as a proxy for representativeness as a whole. You may also assume that non-respondents are more similar to respondents that needed multiple contact attempts than to respondents who agreed to participate immediately. Under this assumption, you might be able to judge whether there is non-response bias if the results of the survey change as harder-to-reach members enter the sample later in the collection process.

However, each of these assumptions are merely assumptions. The best they can do is flag potentially risky cases where you might have to worry, but they cannot tell you definitively that there is or is not a problem.

Moreover, we can only say whether a sample is broadly representative based on traits for which we have data for the population. This is one reason why having a database of member profiles is so useful, since it enables us to say whether a sample is representative based on specific characteristics of interest. This is a major reason why the combination of surveys with a robust CRM is a gold standard for understanding memberships. Without a good set of member profiles, we may instead need to default to industry-wide data and assume that the membership is representative of the industry—something that also would not be directly verifiable.

You might use a sampling strategy that prioritises collecting responses from individuals who are less likely to respond to your surveys. This will require an understanding of patterns of non-response, which can be informed by demographic analysis of past surveys, for instance. You can also consider things such as the mode of your surveys (email, telephone, etc.) and whether a combination of these approaches could be appropriate for reaching different segments of your population. Unfortunately, many efforts to mitigate non-response bias can end up exacerbating it by simply encouraging greater participation from those already inclined to be involved.

For a more in-depth discussion of non-response bias and commonly employed strategies for handling it, see Koch & Blohm (2016).[3]

[3] Koch, A., & Blohm, M. (2016). Nonresponse Bias. GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_004