

# Sampling

Gordon Lynch<sup>i</sup>

## Introduction

One of the aspects of research design often over-looked by researchers doing fieldwork in the study of religion is the issue of sampling. It is relatively common-place for books and articles in the field (particularly written from a humanities perspective) to present their empirical data as being of self-evident value or interest. This is also often an omission from book-length publications as well because thoughtful methodology chapters don't necessarily attract book readers, although discussions of research design and sampling strategy are sometimes included as appendices to a book (see, e.g., Smith & Denton, 2005). As a result much discussion of fieldwork research on religion proceeds with little attention to how researchers chose the sample for their research, or what the implications of this might be for the knowledge generated by their work.

As we will see in this paper, though, questions about sampling can be of fundamental importance to the quality and significance of our research. Understanding different sampling strategies, as well as why we might choose between them in different contexts, is an essential element of thoughtful research design.

There are already some helpful overviews of sampling in textbooks on research methodology (see, e.g., Gray et al, 2007; Bryman, 2008), which discuss ideas covered in this paper in more detail. The aim of this paper, though, is to provide an introduction to some key concepts and approaches to sampling, as well as consider questions about sampling that researchers on religion may often face.

## What is sampling and why do it?

**'A sample** is a proportion or subset of a larger group called a *population*...A good sample is a miniature version of the population of which it is a part – just like it, only smaller.' (Fink, 2003: 1)

This definition of a sample may appear self-evident, but raises fundamental issues about the relationship of our research sample to a wider population. First, how do we define the 'population' in relation to the particular context that we are studying? Sometimes the 'population' will reflect our common-sense understanding of this term in the sense of the population of a particular geographical area (e.g. a town, region or country). But very often, in the study of contemporary religion, such populations do not necessarily map neatly onto a geographical area. Sometimes a 'religious' population may be situated in a specific location (e.g. a religious order or local congregation), but increasingly may take forms that do not map simply onto easily defined spaces (e.g. social movements, on-line networks or people engaged in globalized transactions). Defining the wider population from which a sample is taken therefore requires us to be explicit about what the qualities or traits are that characterize that particular population. How would we know whether

someone does or does not belong to it? It also requires some justification of why these particular defining features are significant. What reason do we have for taking this particular population as the focus? Might our research benefit from re-defining this population in some way? The term 'population' could be misleading in this context if we assume that it necessarily refers to collections of individuals or groups. For field research on religion, a population is quite likely to be a population of particular kinds of events, institutions, texts or objects.

A second issue then concerns how we understand the relationship of the sample for our particular study to that wider population. The nature of that relationship is of fundamental importance to how we conceive the nature and purpose of our research, the kind of knowledge that we claim to generate from it, and the practical procedures we follow in selecting that sample.

One approach is to think about our sample as being statistically representative of that wider population. This means that the findings from our sample could be taken, with varying degrees of confidence, to be representative of that specific population. Such representativeness is also sometimes referred to as the *external validity* of a piece of research, in contrast to the *internal validity* of the suitability of a project's design for examining its core questions. To be able to claim such representativeness requires us to use a *probability* sampling method. This approach is particularly suited for studies in which we are trying to establish generalizable findings for a population whose size can be determined and we can access a sample which is randomly selected from within that population. Because of the statistical requirements for being able to claim such generalizability, this approach requires a rigorous quantitative research design.

Another approach is not to think about our sample as allowing us to generate findings that can be generalized back to its whole population, but to use it as the basis for developing theoretical explanations or accounts which could be applied to that wider population or possibly beyond. This can be referred to as *generalizing to theory* rather than *generalizing to a population*. This approach normally utilises a *non-probability* sampling method, in which the sample is chosen in such a way as to facilitate this process of generating or refining theoretical insights. Non-probability samples are most often used in qualitative research, although quantitative studies may sometimes need to use a non-probability approach in situations where it is not possible to establish a clear sampling frame (see below).

Whilst graduate level research projects would normally utilise one or other of these sampling approaches, it is possible to combine these within a single study. Smith and Denton (2005), for example, made use of a probability sampling approach in order to build a generalizable account of the religious identities and attitudes of American teenagers from a large scale survey. But they also drew on non-probability methods, through their qualitative research interviews, to build up a theoretical account of young people's religious attitudes in terms of 'moral therapeutic deism'.

Fieldwork studies in Religious Studies departments are often qualitative. One of the pitfalls with the presentation and reception of such work is that a non-probability sampling approach is used but the researcher claims, or their audience assumes, that its findings can simply be generalized to a wider population. This confuses the aims and requirements of probability and non-probability sampling methods, and weakens the validity of the way in which findings from such studies are used. Being clear about these different sampling methods, and the different relationships they assume between the sample and population, is important for being more careful about the kind of knowledge we

claim to generate from our work.

### Probability sampling

Probability sampling is used when a researcher is seeking a strong correspondence between their research population and the sample drawn from it. The stronger the correspondence, the greater the degree of 'confidence' (probability) that trends, variations and patterns found in the sample are representative of trends, variations, and patterns that are present in the research population. The stronger the correspondence, the more valid are the generalisations about the research population drawn from the sample.

Probability sampling begins with identifying the *sampling frame*. This is a collection of data that lists all of the constituent units of the population (e.g. an electoral register, a telephone directory, or a list of students registered at a particular university). Identifying an adequate sampling frame is essential for probability sampling as it is from this frame that the subjects of the particular research study will be selected. There is no exact science for knowing that a sampling frame exactly captures all of the population, and given the limitations of any social data, no exact match can ever be guaranteed. The researcher has to come to their own judgment about how reliable the sampling frame is as a summary of the population they are seeking to study.

In an ideal sampling frame:

- all of the units are listed in a way that makes it easy to identify them in some kind of order (e.g. through a numerical identifier), and the frame is organised in a clear and consistent way.
- there is sufficient information about each of the units to be able to access them for the purposes of the study
- every unit in the population to be studied is included, and each unit is included only once
- no unit from outside the population to be studied is included
- there is sufficient information about each unit to be able to make more advanced sampling selections (e.g. deciding to over-sample students with a particular ethnic background)

Care should also be taken to establish whether the way in which units are arranged in the sampling frame in such a way that some methods of random selection might unintentionally lead to bias (or a *sampling error*) in which some kinds of unit were more likely to be selected than others.

When the sampling frame has been established, it is then possible to select the sample for the study from it. There are different ways in which this selection can be made, and more detailed descriptions of these and reasons for choosing between them are provided in Bryman (2008). Some common methods for selection are:

- i) **simple random sampling** (choosing units from the sampling frame randomly, for example, through a lottery, so that each unit has an equal chance of being selected, and there is an equal chance of all different permutations of selections)
- ii) **systematic sampling** (choosing units from the sampling frame by selecting one unit by random and then each subsequent unit at a standard range from that, i.e. every 10<sup>th</sup> unit on the list after the initially chosen unit)

- iii) **stratified random sampling** (defining sub-groups within the wider population and then sampling randomly or systematically within these to ensure that each sub-group is adequately represented in the sample. This approach is helpful when researchers wish to over-sample a particular sub-group within their population, e.g. studying equal numbers of men and women in the sample to compare their responses even though the numbers of men and women may not be equal in the whole population).

By following any of these selection methods, it will be possible to calculate the probability of any individual unit in the population being included in the sample. Knowing this degree of probability is important for being able statistically to assess the degree of confidence in generalizing the findings to the wider population.

When selecting a sample from a population, attention also needs to be given to the size of the sample necessary to ensure a high probability of its representativeness. Obviously the closer the sample size is to the size of the whole population, the greater the probability of it being representative. But with large populations, it is likely to be impractical to sample most of the population, and so reference should be made to statistical calculations of sample sizes required for differing degrees of confidence in its representativeness. This is illustrated in the following table which sets out degrees of confidence of sample sizes for a population of 10,000 units:

Sample Sizes Required for Various Margins of Error, by Confidence Level (Population = 10,000)		
For simple random sampling selection		
Margin of Error +/-	Confidence Level	
	95%	99%
1%	4,899	6,247
2%	2,088	2,938
3%	1,000	1,561
4%	579	942
5%	375	624

[If the population is relatively small (less than 150) then elaborate sampling procedures may not be appropriate]  
(Gray *et al.* 2007: 113)

Sometimes researchers, or consumers of their research, suggest that the generalizability of findings from a particular study can be read as a general statement on society. This can often happen, as well, when striking findings from a particular study are repeatedly discussed in conferences or other publications but without close reference to the specific population from which the study drew. A more rigorous view, though, is that the generalizability of findings from a probability sample can be generalized only to the population captured by the sampling frame. If our sampling frame, for example, only includes clergy from a specific range of Christian denominations in a particular country, we cannot assume that findings from the study can be generalized to all Christian clergy in that country, or indeed to religious workers across a wider range of religious traditions. Knowing the

sampling frame for a probability sample is therefore important for being able to evaluate how findings from that study might be applied.

### Non-probability sampling

Non-probability sampling methods aim to construct a sample that can generate the most useful insights that can be gained by the researcher into that study's particular focus. There are a variety of reasons why it may not be possible or desirable to establish findings that are generalizable to a wider population. Sometimes it might be important to establish insights about a group or phenomenon for which we don't have the data to create a rigorous sampling frame. We might want to use a survey, for instance, to understand the attitudes, experiences and practices of Sikh university students in the UK without having any clear data about how many Sikh students there actually are in the UK higher education system (because the religious identifications of UK university students are not measured at a national or institutional level). This does not mean that the findings of a survey based on a non-probability approach are necessarily invalid, but that we just use them differently to findings based on a probability sample.

Alternatively we may wish to focus our research on an in-depth study on a smaller number of cases or participants to allow us to develop more complex insights. This may involve studying people or events who are not representative of a wider population, but who are worth studying precisely because they represent exceptional, critical or intense examples of the particular phenomenon we are interested in. There may be particular sources of data which the researcher is able to access in a way that is unusually rich, and which could not be obtained in the same way by other researchers. Their findings are not, therefore, straightforwardly generalizable to a wider population, but nevertheless have the potential to generate valuable insights. The basic aim here, though, is to be able to say something *theoretically* about the findings from the particular sample we have studied in a way that generates insights or questions about other cases or contexts. One implication of this is that studies based on non-probability samples need to be designed in a way that is consciously informed either by existing theoretical debates or the development of new theory using an approach such as grounded theory.

A good example of a theoretically-useful study using a non-probability sample is Robert Orsi's use of studies of members of his extended family in the context of a history of twentieth-century American Catholicism in his book *Between Heaven and Earth* (Orsi, 2005). Whilst Orsi's use of data from his own family history clearly raises interesting methodological questions, it also means that the duration and complexity of his contact with the people he writes about would never normally be possible for most researchers. When he writes in depth, for example, about the life of his Uncle Sal, who was born with cerebral palsy, Orsi does not present Sal as statistically representative of disabled American Catholics in the twentieth-century. But this does not mean that his account of his life lacks useful broader insights. Through the careful tracing of this life, Orsi is able to show the complex ways in which embodiment, family relations, intimacy and emotional distance, space, broader institutional and historical structures, and inter-subjective relations with sacred figures intersect in the ways in which the subjectivity of disabled Catholics was experienced, imagined and contested. In one sense, Orsi's narrative about his Uncle Sal tells us primarily about his uncle. But by situating Uncle Sal's life in a broader theoretical understanding of human subjectivity and social structure, Orsi is able to

write about his life in a way that touches upon these broader aspects of social life. Furthermore, by situating Uncle Sal's life in a broader historical account of perceptions of disability in popular Catholic media and literature (a form of *triangulation* of data), Orsi is able to show how elements of Sal's experience are replicated in other lives and contexts beyond his individual life. This does not mean that insights about Sal's life are statistically generalizable to a wider population. It is not possible to say, for example, that a certain percentage of disabled American Catholics had the same experience as Sal, but then research of this kind suggests that human experience always weaves together the idiosyncratic and wider social patterns and that straightforward generalization about richly-understood experience is not possible. A wider observation can be drawn from this. There is a sense in which the findings from a study based on a non-probability sample apply most directly to those or that which is included in the specific sample. But it would be fallacious to argue that the findings are only relevant to that specific sample. If a non-probability sample is selected and studied in theoretically thoughtful ways then this should generate insights about lives and contexts which involve similar social or cultural dimensions or in which a comparable theoretical approach is being used. Problems arise, however, when studies use non-probability samples without any relationship to theoretical concepts that make it possible to relate the sample studied to broader social processes, and findings are then reported with the implication that they are generalizable. Non-probability samples do not yield generalizable findings, but ways of seeing data that may be applicable to other cases.

Within a non-probability approach to sampling, a range of different methods are used by researchers. These are sometimes also referred to as *purposive* sampling techniques, because the researcher is intentionally choosing the size and content of their sample in order to maximise their learning from their project. In making these selections, researchers may be choosing not only which people to include in their study, but what times, places, events or interactions. Most common amongst these methods are:

- i) **theoretical sampling:** the sample is chosen with the explicit intention of examining existing theoretical insights or developing new ones. In the case of the former, participants may be selected on the basis of how well they enable key elements of that theory to be explored or tested. In the case of the latter, on-going choices may be made about who or what to include in the sample in order to check and refine emerging theoretical insights. Within this approach to selection, researchers may also opt for a *maximum-variation* sample which includes a range of differences in factors that are taken to be important for the phenomenon under study or a *homogenous* sample in which participants are chosen primarily because they reflect a common set of traits or factors.
- ii) **convenience sampling:** the sample is selected primarily on the basis of what the researcher is able to access. Whilst this is often a default approach in small-scale pieces of research (e.g. undergraduate or Masters' dissertations which may rely on the writer's existing contacts), one of the strongest rationales for this method is when the group or phenomenon under study is generally difficult to access but the researcher is able to establish a sufficient degree of contact or trust with particular participants to conduct a viable project.
- iii) **snowball sampling:** the researcher builds their sample on the basis of contacts suggested by other participants. This potentially has the advantage of drawing on

participants' own expertise in developing the sample as well as expanding the sample beyond contacts known to the researcher in the first stage of their project.

In practice, non-probability sampling decisions are likely to be influenced by a range of judgments by the researcher about the degree of co-operation and quality of data that they would get from different sources, and the relevance of that data to the key, theoretically-informed, questions that they are trying to answer.

### How many is enough?

A common question in research supervision and methods classes is 'how big should my sample be?' From the discussion so far, it will be clear that the answer to this depends initially on whether we are trying to construct a probability or non-probability sample. In the case of a probability sample, the size of the sample needs to be such that it gives reasonable confidence of generalizability to the wider population. In this sense, the answer to this question is a statistical one. It depends on the size of the wider population, and what sample size is needed to achieve particular degrees of confidence about its representativeness (as illustrated from the table from Gray et al (2007) given above).

In the case of non-probability sampling, there is no 'correct' number of participants. Instead it is more useful to think in terms of the richness of the data in terms of answering (or refining) the researcher's core questions. It would be possible, for example, to produce a successful PhD thesis based primarily on the case of a single person. But this would require a highly sophisticated understanding of why that individual case was significant (including the ability to contextualise it theoretically, historically and socially), what we could learn theoretically from that single case and a range of data from that person that was characterised both by substantial depth and extent. Nevertheless, a theoretically-grounded and richly-informed case of a single person would be significantly preferable to a study involving thirty participants which lacked any theoretical-grounding or social and historical contextualization and was based on superficial data. The number of research participants in a non-probability sample is not, therefore, a marker of the quality of the study. Rather what matters is the researcher's ability to access people and data that enable them to answer their research questions as insightfully as possible. Often these judgments will develop as the project itself is underway, as the researcher becomes clearer about lines of enquiry or theoretical lines of reflection to pursue. In this sense, starting with a 'target number' of participants may be unhelpful in the sense that it suggests that the target number is any guarantor of quality for the research findings. Focusing on what is needed to answer the question is more valuable.

A common suggestion for qualitative, non-probability studies is that researchers should continue to develop their sample until they have reached 'saturation' in their data (i.e. when researchers are not generating any significant new insights in relation to their core questions). In some cases, this can be helpful, although in others this can lead to researchers focusing on certain common themes as a way of providing some kind of neat closure to their project rather than continuing to attend to the complexities and contingencies of their data. Another way of thinking about this, which can be both more pragmatic and theoretically and contextually sensitive, is to recognise that the process of data collection is always limited by time and other resources. For most doctoral students, there is little opportunity to spend much more than a year of their period of registration engaged in direct

fieldwork (with the rest of their registration taken up with developing their literature review, formulating their research questions and writing up). This means that the sample may never be 'complete', but it invites the researcher to think critically about the implications of what sources of data they have, or have not, been able to access (which in itself can be highly informative about their field of study), and to be explicit with their reader about the ways in which the eventual sample shares certain commonalities with other contexts and in what ways it is distinct.

### Comparing probability and non-probability sampling

Two misconceptions that can be held about non-probability samples is that they are less rigorous than probability samples, or that they are just as good as probability samples if a researcher doesn't have the training, inclination or resources to do a quantitative study. Both misconceptions are wrong because they fail to recognise that constructing probability and non-probability samples are different ways of designing research which produce different kinds of knowledge-claims, and whose appropriateness relates to the particular aims of the study in which they are used. A study based on a non-probability sample can be rigorous, but its rigour and the kind of knowledge it generates are not the same as for a study based on a probability sample.

A more constructive way of thinking about this is in terms of the relative strengths and weaknesses of well-designed research based on probability and non-probability samples. Probability samples obviously enable us to make inferences about wider populations. Given the interest, for example, in religious trends in modern society such as secularization and desecularization, the effects of migration on religious commitment, changing social values or the rise of religious 'fundamentalisms', clarifying such patterns in rigorous and measurable ways is clearly important. This is arguably particularly important in the study of religion where theoretical assumptions about religion in the modern world have not always been subject to rigorous measurable studies. When such generalizable research is done it does not always turn up results that reflect previous academic assumptions. One of the most striking findings, for example, of Heelas and Woodhead's (2005) study of the 'holistic milieu' in the town of Kendal, was how small a percentage of the population (less than 1%) took part in activities within this 'holistic milieu' which they considered to be of spiritual significance. This gave a helpful perspective on the assumption that 'New Age' spirituality is a widespread feature of contemporary society. Measuring carefully, and generalizing on the basis of probability sampling, can be an essential check for ideas that have become a consensus amongst scholars without a strong empirical grounding.

At the same time, if we restricted ourselves only to research using probability samples this would narrow our scholarship far too much, both in terms of the groups we might study and the kinds of questions we might explore. The measurement of larger numbers of research participants, typical of probability samples, inevitably imposes limitations on the depth and complexity of the data we are able to collect. We may be able to perform complex statistical operations to establish different kinds of patterns within that data, but the categories we use to collect that data inevitably need to be sufficiently simple to be written into straightforward survey questions or other standardized measurements. Generalizable studies using probability samples may, for example, be able to establish levels and salience of 'religious belief' in a population, but be less able to tell us how applicable conventional notions of 'religious belief' are to people's everyday lives, at what moments



or contexts in their lives different forms of belief become more or less relevant, or the relational context that shapes the form that 'belief' takes in their lives. To understand such contextual issues of 'belief' in a more sophisticated way, we will normally need to adopt a qualitative approach, based on a non-probability sample. Doing this can help to clarify various insights and assumptions that we bring to the category of 'belief', which might then inform the way in which we interpret data from generalizable studies based on probability sampling.

Research based on probability samples should not therefore be seen as intrinsically more valuable than that based on non-probability samples, and vice versa. Rather they should be seen as different approaches to research design which help to create different kinds of knowledge, and which can have a complementary rather than conflictual relationship in helping us to understand better our social worlds.

### A messy postscript

One of the positive uses of the concepts of sampling described above is that they can give a degree of clarity to researchers (particularly those developing their first substantial projects), about what they are doing and why they are doing it. They also establish a set of conventions for what we might reasonably claim, or not claim, for the findings from our research, and help us to think about the basic aims that inform our projects.

There can also be certain risks with such methodological clarity. An implication can be that if we follow these conventional procedures, then the social world will yield its knowledge to us. But this can imply that our methods (including our sampling methods) are neutral means to accessing knowledge, when in reality research methods play a role in constructing the social worlds that they study. As John Law (2004) has argued, social realities always exceed our capacities to know and represent them. As a consequence:

'Method is not... a more or less successful set of procedures for reporting on a given reality. Rather it is performative. It helps to produce realities. It does not do so freely and at a whim. There is a hinterland of realities, of manifest absences and Othernesses, resonances, and patterns of one kind or another, already being enacted and it cannot ignore these. At the same time, however, it is also creative. It re-works and re-bundles these and as it does so re-crafts realities and creates new versions of the world. It makes new signals and new resonances, new manifestations and new concealments, and it does so continuously. Enactments and the realities that they produce do not automatically stay in place. Instead, they are made, and remade. This means that they can, at least in principle, be remade in other ways.' (ibid., p.143)

Understanding sampling procedures is invaluable for learning the conventional craft of what it means to be a social researcher. But this should never shut down our awareness of the complexity of the relationship between our research and the social worlds that we study. Questions of ontology, epistemology and the politics of research always need to be kept in mind as the bigger picture within which our thinking about sampling is framed.

## References

- Bryman, A. (2008) *Social Research Methods*. 3<sup>rd</sup> edition. Oxford: Oxford University Press.
- Fink, A. (2003) *How to Sample in Surveys*. 2<sup>nd</sup> Edition. Thousand Oaks: Sage.
- Gray, P. S. et al. (2007) *The Research Imagination: An Introduction to Qualitative and Quantitative Methods*. Cambridge: Cambridge University Press.
- Heelas, P. & Woodhead, L. (2005) *The Spiritual Revolution: Why Religion is Giving Way to Spirituality*. Oxford: Blackwell.
- Law, J. (2004) *After Method: Mess in Social Science Research*. London: Routledge.
- Orsi, R. (2005) *Between Heaven and Earth: The Religious Worlds People Make and the Scholars Who Study Them*. Princeton, NJ: Princeton University Press.
- Smith, C & Denton, M. (2005) *Soul Searching: The Religious Lives of American Teenagers*. New York: Oxford University Press.

---

<sup>i</sup> I'm very grateful to Andy Dawson for the original development of training material on sampling for the Oxford seminar which informed the construction of this paper.