Researching religion using quantitative methods

Ingrid Storm
Institute for Social Change, University of Manchester

Quantitative research is in principle anything that involves classification or measurement. A considerable amount of the research conducted on religion and religiosity is implicitly quantitative by addressing questions about the prevalence of religious practices, the growth or decline of religious movements or the intensity of religious beliefs. Measuring religious variables lets us say who is religious and to what extent, by age, sex, ethnicity, marital status, class and so forth. It can also be used to explore how religion and religiosity are associated with values, attitudes and behaviour, and allows us to test theories about the causes and consequences of religious involvement.

Whether or not you have used quantitative data and methods before, this short introduction will give you a taste both for how the topic of religion lends itself to this type of research, and also how you can apply quantitative methods to answer your own research questions.

1) Why quantitative data are appropriate and important to the study religion?

In general quantitative methods have the advantage of allowing relatively precise hypothesis testing, and are less vulnerable than qualitative methods to theoretical bias in the data (Brink 1995). However, there are objections to studying religion using quantitative data. Three of the most common ones are 1) that religion is too complex to be classified and measured at all, 2) that quantitative methods are too simplifying and empiricist to be used within a non-positivist epistemological framework, and 3) that religiosity is too context-dependent and sensitive to measurement error to be reliably quantified. Each of these will be addressed in turn, before turning to more practical advice about different forms of quantitative methods and how they concretely can contribute to the study of religion.

One response to the first of these objections, that the phenomenon of religion is too complex to attempt to quantify at all, is that the effort to classify and measure forces us to be clear and open to criticism. While it is true that religion is a complex phenomenon, this makes it all the more important that researchers are clear about which aspects of religion their analyses are concerned with and to what populations their results can be generalised to.

Moreover, regardless of methodological preferences, some questions are simply better suited to be analysed by quantitative methods, or may even be dependent on them. For example, hypotheses about the alleged growth in alternative spirituality, the apparent strength of evangelical or charismatic congregations, the relative religiosity of women, and so on, are impossible to answer without quantification and measurement of some kind. In short, quantitatively framed questions demand quantitative answers generated by quantitative methods. While qualitative research could add to knowledge about these hypotheses, such as provide a richer understanding of the mechanisms of growth
of a religious movement, they cannot by themselves test whether the growth is taking place.

While quantitative methods are sometimes the best suited to answer a research question, it is important to recognise that these methods are just as in any other method, influenced by the researchers’ choices of what to analyse and how to analyse it. As Swatos (1977), puts it “Quantification has no special magic because it does not in itself "unbias" research”. In other words, rigour is not a given in any kind of research. However, quantitative methods has the advantage that it makes it possible for a good researcher to honestly pursue the goal of rigour and to detail the use of methods so that the research can be replicated, tested and criticised by other researchers in the scientific community.

2) Epistemological and methodological perspectives

It is an unfortunate, but influential myth that all quantitative social research is based on a positivist approach to knowledge. If this were true it would be a hindrance to the use of quantitative methods for most researchers of religion, but fortunately it is not the case. The quantitative study of religion is conducted from a number of epistemological perspectives ranging from positivism to constructivism. That being said, most academic research is founded upon a basic ontological belief that there is an observable reality external to the researcher, and the quantitative study of religion is no exception. The approach labelled critical realism (Bhaskar 1997) is probably the most common in quantitative research on religion. It holds that there is an external reality, and that knowledge about it can be approximated through research, while at the same time rejecting the epistemic fallacy that reality can be reduced to its observable representations (Danermark et. al. 2002:39). More fundamentally, one assumes as a researcher of religion (or any other human experiences, concepts and abstractions) that reality cannot be reduced to either matter or mentality (Russell 1921). Any experience depends on a relationship between physical matter and a conscious mind, and neither would on its own do justice to the phenomenon under consideration. The most relevant feature to focus ones study upon is thus neither the observable (“how many people can be seen in a particular church on a given Sunday?”), nor the purely mental (“do these people think God exists?”), but the experiential relationship the people in Church have with each other, with the observable symbols and images they encounter in the church, and with unobservable deity they are worshipping. While religions are socially constructed, they are also subjectively lived. It does not make the object of study any less “real”, to acknowledge that it is shaped by a combination of language, cultural tradition, physical objects, environmental factors, genetic traits, bodily practice and individual mentalities (etc.) and is irreducible to any of these. However, ones emphasis might affect the way one observes and measures, and most importantly analyses it, and awareness of this is crucial. It is naturally a challenge to the researcher to keep everything in mind at once, but it is a challenge that can best be met by interdisciplinary openness to different methods and scholarly traditions.

The use of quantitative and qualitative methods respectively is an important methodological consideration for any project (Bhaskar 1997). Despite their common association, there is no necessary relationship between a realist paradigm and an
empiricist quantitative method (Creswell and Tashakkori 2007:304; Bazeley, 2002:4), nor between a social constructivist approach and qualitative methods. Constructivist researchers may be more interested in the co-construction of knowledge between researcher and researched than with the knowledge itself, but there is no obvious reason why this process of construction cannot be studied through the use of quantitative methods. Anyone designing surveys knows that there is plenty to be studied. How one formulates a question, what part of the questionnaire it is placed in, whether it is asked over phone or in person, and who is asking it are all potential influences on how the participant answers, and consequently how this answer is analysed. In other words, the process of knowledge-construction can be and is being studied using quantitative methods. For example Deborah Glik (1990) uses a combination of survey data and interviews to argue that spiritual healing processes are socially constructed events. Due to the subjective and constructed nature of the personal identifications, beliefs and attitudes that constitute religion, it is considered highly appropriate to also approach this through qualitative methods such as interviews. However, when issues of representativeness and generalisability are at stake, a quantitative approach may be more appropriate. Even strictly “qualitative” researchers may wish to use quantitative methods as a supplement to their research, for example as background information to evaluate the representativeness of the findings, or as an analytical technique for content analysis of qualitative data. Hence it is important to understand the advantages, but also the pitfalls of studying religion with mixed methods.

According to Flick (1992) mixed method should be employed, not as a strategy for validation, but rather as a means of getting “access to different versions of the phenomenon that is studied” (Flick 1992:194), while ensuring that the methods and criteria are appropriate to the subject (Flick 1992:175). Because the use of different methodologies may hinder the direct comparability of data, mixed methods may be better used for achieving complimentarity, contextualization and elaboration rather than corroboration of results (Bazeley 2002:4). Through the approach of “intersubjectivity” (Morgan 2007:71) the pragmatic researcher can work back and forth through different methods and achieve mutual understanding between different frames of reference. This may be particularly useful in studies where the internal experience of individuals is understood in reference to the wider social environment, as Christian Smith’s (2005) study of the religiosity of American teenagers demonstrates. He analyzed large-scale quantitative survey research to present an overview over the religious landscape of teenagers in America. However, he used personal interviews with a sub-sample in order to explore the complexity and contradictions in individuals’ religious beliefs, values and identities. In other words, this is not an example of integrated mixed methods, since the methods are strictly speaking not used to study exactly the “same thing”. Brown’s (2001) The Death of Christian Britain is another example of a historical study which relies on a combination of on qualitative textual analysis and quantitative data from a number of sources to explain the causal processes in the secularisation of Britain. These pragmatic approaches let the research question guide the methodology, while recognizing the limited comparability of results from different methods.
3) Problem of measurement

The problem of measurement and reliability is probably the most valid critique of quantitative studies of religion. Survey research on religion does seem to be particularly sensitive to context, wording and expectations, and should therefore be approached with caution. For example, if you ask people to identify their religion, the answers may be very different depending on the exact wording and context. Table 1 illustrates this with the results from the 2001 Census in England and Wales compared to the British Social Attitudes survey from the same year:

<table>
<thead>
<tr>
<th></th>
<th>Census</th>
<th>BSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christian</td>
<td>71.7</td>
<td>54.2</td>
</tr>
<tr>
<td>Non-Christian</td>
<td>5.7</td>
<td>4.7</td>
</tr>
<tr>
<td>No religion</td>
<td>14.8</td>
<td>40.5</td>
</tr>
<tr>
<td>Not stated</td>
<td>7.7</td>
<td>0.6</td>
</tr>
</tbody>
</table>

The difference between the question wording here was that in the Census, to reduce the length of the questionnaire, the respondents were simply asked “What is your religion?” with tick-box options that listed world religions. In the BSA the question was asked in two steps, the first being “Do you belong to a religion? (Yes, or No)”, the second asking those who answered yes to the first question, to specify “Which one?”. An additional difference is that because everyone has to fill out the census forms, they are often completed by the household head on behalf of all individuals at the address. Since the person doing this tends to be older and more religious than average, the numbers may be higher than they would be on confidential individual questionnaires.

In addition to the wording of the question, the context matters too. One important contextual influence a questionnaire is the order of the questions. This may explain why according to the 2001 census, fewer people in Scotland had a religious affiliation compared to England and Wales, despite coming out as more religiously affiliated in most other surveys. This map illustrates the results: in the darkest-coloured districts more than 25% of the population responded “No religion” to the census question.
There were three important differences between the Scottish census and the English and Welsh one. Firstly, the religion question used on the census form in Scotland came before (rather than after) the question on ethnicity, secondly it was worded in a less leading way and thirdly it also offered answer categories for specific Christian denominations. Perhaps as a result, people were nearly twice as likely as in England to give their affiliation as `none'.

This example also highlights another important issue concerning the comparability of data across countries and religious traditions. While many surveys are designed to be internationally comparable, slight differences in language and administration of the questionnaire could potentially have an impact on the results. A further potential criticism of a cross-national comparative study of religion is that while religious behaviour may look similar in the survey data, the meaning of that behaviour may vary drastically depending on the cultural context. For example, weekly churchgoing is a much more conventional behaviour in Ireland than it is in England. In some cases these national variations in meaning are precisely the objects of study. In other cases some steps could be taken to ensure as much comparability as possible, for example by comparing religiosity relative to the average for each country, rather than uncritically comparing like for like.

Needless to say, the purpose of mentioning these issues is not to discourage use of quantitative data. However, it is a cautionary note about the importance of questionnaire design. It is also an illustration of how the corroboration of multiple sources of data can alert one to possible systematic problems with the data collection methods, whether it be issues of sampling, questionnaire design or simply a slip-up somewhere in the process.

4) A beginner’s guide to researching religion using quantitative data

For researchers of religion who wish to use quantitative data for the first time there are a number of issues to consider. Firstly, one needs to be aware of the sources of quantitative data. While most of the examples mentioned here are based on large scale national surveys, there are a number of other sources of quantitative data on religion, such as administrative records (e.g. church records) and commercial statistics (e.g. book sales). It is also possible to obtain data through the coding of qualitative data, such as media time / counts, textual content analysis and spatial analysis. Some data such as church records or large scale surveys are obtainable as secondary data, whereas others you can collect yourself, either by administering a questionnaire, or by counting (e.g. number of people who enter a mosque on a Friday, the number of times “God” is mentioned in a radio show etc.).

Data collected by the researcher is called “primary data”. The advantage of primary data gathering is that you can tailor-make your own survey questions, research site and sampling to suit your particular research questions. The downside is that it is often time consuming and potentially expensive. It may also be difficult to achieve a sample that is sufficiently large or representative. In contrast, with secondary quantitative analysis, the data is immediately available, and often of high quality. National government-funded surveys typically have large random samples and cost
millions of pounds to produce. However, if using one of these surveys one only has answers to the questions asked and so it is important to make sure that the information suitable to answer your research question is available before picking a secondary data source.

It is also important to differentiate between aggregate statistics, such as numbers in tabular form or opinion polls that are based on analysis by other people, and microdata, that is the raw data records giving items of information (variables) for each individual (cases). Webpages such as BRIN and ARDA are excellent sources for the former, but to produce your own charts and tables you need the original microdata. The Economic and Social Data service (ESDS) is a good source for British survey data, and their website has a search tool for variables, where you can enter your topic of interest and see which surveys have covered it. This could also be useful if your are planning to collect or if you have collected your own data and you want to compare any of the questions and distributions with those of a representative survey.

**Cases and variables**

A very important consideration for all quantitative analysis is defining the unit of analysis. In quantitative research a “case” is the entity for which we have data and while this is usually an individual person in social research, it could also be a mosque, a neighbourhood, an event, or something else. Each case is measured by the same variables.

A variable is, as the name indicates, simply something that varies, and used in statistics it is something you can measure the variability of. An important distinction to keep in mind when using quantitative data is that what you want to measure may be either discrete (in which case you count them: more or fewer) or continuous (in which case you use some yardstick: more or less) Sometimes they also have an order (higher / lower, bigger / smaller, better / worse).

The difference is important for how we can quantify, measure and analyse the variables. Consider the difference between “Religion” and “Religiosity”. Those variables that operate with discrete categories, are called “categorical variables”. These are measured by categorising. For example Religion can be categorised as “Christian”, “Muslim”, “Buddhist” and other. Categories can be either very broad or quite specific, but they are rarely irreducible. For example the category Christian, can be further narrowed down by defining in terms of other attributes such as Bible reading, orthodoxy, etc. The important thing is that the categories are exhaustive, (all possible outcomes fall into one of the categories) and exclusive (no outcomes fall into more than one category). In practice this may require some judgment on the part of the researcher. What do you do with someone who identifies as Buddhist and Christian for example?

Most categorical variables are nominal variables. This means that there is no particular order to the categories. When it does matter what order the categories are in, we call them ordinal variables. An example would be socioeconomic classifications, or ordered responses to a survey question such as “How religious would you say you are? – “Very”, “not very” or “not at all”. In the same way that it is difficult to quantify
the difference between middle class and working class, in this example, we know that the response “very” is *more than* “not very”, but we cannot say *how much more*.

Quantifying **continuous variables** like “religiosity” or “numbers of members in a congregation” is often done implicitly. Every time words like ‘more’ or ‘less’, are used it implies quantification and comparison. If the quantification is to be made explicit however, we need a *scale* to measure it. With some variables this is quite straightforward - we can simply count how many people are sitting in the pews of a church, or look at the how much money was donated to a particular Muslim charity organisation last year, or ask someone how many miles they walked on their last pilgrimage to Santiago de Compostela. The point is that we know not only that 500 miles is more than 100 miles, but we can also say *how much* more. This is what distinguishes continuous from ordinal variables, and this has consequences for what statistics can be used for analysis.

With other variables such as frequency of prayer it is more complicated – while we can count, by asking or observing how many times the respondent prays within a set time period, for example a week or a day, this is not without problems. Firstly designing such items often depends on assumptions on how often we would expect people to pray in the population we are studying. More fundamentally, there may be a qualitative difference between people who never pray and those who pray once a day that is incomparable to the difference between those who pray once a day and twice a day.

With something more abstract like “degree of religious commitment” it is impossible to count it directly. However, it is possible to measure it as a continuous variable. One option is to ask respondents to place themselves on a scale, but the problem with this is that it is difficult to determine whether all respondents have the same criteria for what counts as “very religious” etc. Another option is to combine a number of ordinal variables on various religious beliefs and practices into a single scale, or by otherwise identifying quite precisely how religious an individual is. This is where the importance of conceptual definition and clarity becomes evident.

**Sampling**

Because its purpose is to generalise, sampling is fundamental to quantitative research. There are two crucial points about quantitative samples: their size and their random selection. Firstly it is important to ensure that the samples are sufficiently large. Interviewing 20 people from different age groups, genders and social class who seem to be a cross-section of the nation is no guarantee they will also be representative in terms of their religiosity or values. They *may* be representative of the population as a whole on these variables too, but the problem is that we have no way of knowing. Secondly it is important that the samples are random. Having 2000 people answer a survey on a webpage is no better than having 20 if there is no way of telling what they represent. There might be inherent selection bias in the type of people who would visit the webpage. Even if the 2000 people are selected randomly, there is no guarantee that they will be representative of their population. The difference is that we can calculate exactly how likely it is that their responses will not be representative.
With a large random sample, the probability of generalising successfully to the population is very high.

**Descriptive and inferential statistics**

So far we have treated the “quantitative approach” as if it were a single method, but there are a number of different ways of analysing quantitative data. An important distinction to make is between descriptive data analysis on the one hand and inferential analysis on the other. Generally, good and interesting academic research tries to do both. We want to make sure we know what is happening in the data before we can attempt to explain it or extrapolate it to the population. However, only describing a phenomenon in the data does not necessarily add to our understanding if we cannot also say whether it can be generalised to the population, and perhaps even something about the reasons and mechanisms behind what we observe.

Example of descriptive statistics measures are the mean (average), the variance, and distributions. These are useful for summarising data and presenting in a manageable form and can be expressed either as numbers, or graphically in a chart. *Univariate* analysis is looking at one variable at a time. For example religious affiliation to see how many Catholics, Hindus etc. are in the sample. Descriptive analysis can thus give the researcher a sense of the sample as a whole, but it is also useful for comparing different groups within the sample. For example, one might want to know the percentage of Muslims in different countries. This means we have to introduce a second variable, country, into the analysis of religious affiliation. Analysis with two variables is called *bivariate* analysis. An example of a study based primarily on descriptive analysis is William M. Newman and Peter L. Halvorson’s (1979) thorough quantitative analysis and mapping of the geographic distribution and migration of *American Jews*.

Inferential statistics are methods which allow us to test hypotheses about the population using the sample data. In order to do this one must test that the findings from the sample are not merely due to chance, but are likely to hold for the entire population. Statistical significance means simply that the results are unlikely to appear by chance. By convention, our minimum threshold is a 5% chance that they would arise in the absence of a real effect. If the findings are very strong, they might be described as ‘highly significant’, which means that the probability of a false effect is less than one in a hundred. In other words, that a finding is “significant” does not mean that it is interesting, nor does an interesting descriptive analysis of the data necessarily represent a “statistically significant” result.

Even after testing for significance using a statistical test, there is still no guarantee that we know what the finding means. A typical situation in bivariate analysis is that we are left wondering whether a statistically significant relationship is a direct association between two variables or is due to some other influence. For example, if we find that people who are religious are less likely to have higher education than those who are nonreligious, we might wonder whether this is because education affects religiosity or vice versa – but there is also the possibility that a third variable affect both of religiosity and education. In this example, knowing that younger people
are more likely to have higher education and also less likely to be religious, we might want to control for age.

**Multivariate regression**

If quantitative analysis is the primary research method, then multivariate analysis (analysis using more than one variable at a time) is often regarded as more interesting, since it can tell us more about the mechanisms by which an association occurs. The most common form of multivariate analysis is regression models. Regression analysis has the advantage that variables can be introduced in steps, one by one (or several at once), enabling comparison between simple and more complex models. Generally, one would choose to control for variables that might be relevant to the outcome without being of main interest. Hence, most regression models include sociodemographic variables such as age, gender, education etc. What else one might want to control for depends on the particular research question. If the research concerned whether religiosity is associated with knowing your neighbours for example, it might be relevant to control for type of housing, type of area (urban, rural etc.) and whether the respondent has moved recently as these are variables that are potentially important to the outcome variable.

Selecting which variables to include is chiefly dependent on theory, but ideally it is also based on previous research. Even more fundamentally however, it depends on the availability of relevant variables, with appropriate categories. Once these variables are included we can measure how much of the variation in neighbourliness is explained by these background variables, and how much, if anything, religiosity adds to the model. If the model is really good we can also predict the outcome based on a specification of the independent variables. To go with our previous example we can estimate about how many neighbours someone will know if they are a 62 year old Christian churchgoing woman living in a semi-detached house in a suburban area. Multilevel regression models can also control for individual and context variables separately. This is especially useful for cross-national research where differences between countries or regions may account for substantial part of the variance between individuals. An example of a multilevel model is Barrett et al.’s (2007) study of the impact of school environment on adolescent religiosity, employing both school level and individual level variables.

**Longitudinal analysis**

One of the main reasons for using quantitative data in the study of religion is that it is necessary for testing hypotheses about the growth and decline of religious beliefs, practices or movements. However, in order to systematically address questions about change over time, one needs access to data covering a series of time points. There are two forms of longitudinal analysis that are relevant for the study of religion. The first is trend studies. This form of longitudinal analysis does not necessarily require longitudinal data measuring exactly the same respondents or cases at different time points. All it requires is the same data from the same population at different time points.
Hence if one were interested in measuring the developments in belief in God in Britain, one could use something like the British Social Attitudes survey. The data itself is cross-sectional, that is different respondents are asked each year, but because the selected respondents are representative of the population we are interested in (the population of Britain), we can still measure aggregate trends in the belief in God. We can see from such data whether more or less or the same amount of people in Britain believe in God now compared to thirty years ago and we can control for some factors that may explain the difference, such as different levels of education, age composition etc. Looking at the age of the individuals can give us an impression of whether the change is due to an age effect (change in the beliefs of individual respondents) or cohort effect (change resulting from the change in population, for example if more believers have died than have come of age during the period).

What cross sectional data cannot do however, is tell us anything why individual respondents might have changed their mind about the existence of God during the thirty year period we are looking at. Using longitudinal data such as Panel studies enables us to follow the same respondents over time, analysing how and when they change their responses. This sort of analysis enables us to test hypotheses about the reasons and mechanisms of change as well as describing the change.

Voas and Crockett’s (2005) paper makes use of both trend analysis and longitudinal panel analysis, describing aggregate change over time and analysing change within the same individuals over time. It also includes a useful introduction to Age, period and cohort effects.

Data reduction techniques

A crucial issue from the outset of rigorous empirical investigation has concerned how best to ‘operationalise’ religion and religious commitment. Two scholars have been particularly influential in this discussion. Charles Glock (1959) defined the core dimensions in religiosity as belief, knowledge, experience, practice (and consequences).

Gerhard Lenski (1961) identified two dimensions relating to personal religiosity (doctrinal orthodoxy and devotionalism) and two to types of interpersonal involvement: associational (within the institutional context) and communal (social interaction outside the church setting).

In other words, religiosity is a complex phenomenon, and consists of a variety of different beliefs and practices that may vary considerably between each other. Nevertheless by using the concept religion we assume that all its dimensions, including belief, practice, affiliation and so on, are in some way indicative of a singular characteristic of the individual, which we may call “religiosity”. One way of measuring this assumed underlying attribute, is to create an index of religiosity.

The rationale for creating a multi-item index of religiosity is twofold: firstly it is a way to measure this assumed latent variable of religiosity that each observed variable, such as belief in God or frequency of Church attendance may be indicative of. Secondly it is a way of increasing reliability and reducing the random error and
inconsistencies that arise from using the particular observed variables on their own. By combining several variables the errors balance out and the index becomes more reliable than each of the variables on their own (Voas 2009: 419). However, there is possibility that the multidimensionality is so great that such a construct does not reflect a single underlying attribute. There are a few analytical methods that can help us discover whether or not this is the case, and how to best construct an index that reflects the data. Factor analysis is the most used method. The main applications of factor analytical techniques are firstly to reduce the number of variables and secondly to detect structure in the relationships between variables, that is to classify variables according to how they relate to each other. For example one could imagine a situation where there is not just one dimension of religiosity, but where different individuals score very differently on three different dimensions of religiosity. Let’s call them the “spirituality” dimension, the “belonging” dimension and the “religious activity” dimension. Factor analysis helps the researcher to identify different “dimensions” of the data, and these can subsequently be used to create reliable indexes. Factor analysis is particularly common in psychological research, and one example of this is Rowatt and Kirkpatrick’s analysis of attachment to God:

There are alternative data reduction techniques such as latent class analysis or cluster analysis, which rather than identifying dimensions of variables that are related, identifies different groupings of respondents depending on how they score on a set of chosen items. An example of cluster analysis can be found in Storm’s analysis of different forms of “fuzzy fidelity” (being neither religious nor nonreligious):

Bibliography

Brink, T.L. 1995. 'Quantitative and/or qualitative methods in the scientific study of religion'. Zygon 30: 461–475.


