PART I

Online Quantitative Survey Research

The single biggest impact of the internet on market research to date has been on the collection of quantitative survey data. The chapters in this first part of the book address different aspects of quantitative survey research conducted via online surveys.

Part I covers

- Overview of quantitative online research
- Web survey systems
- Designing online surveys
- Online access panels
- Client databases
- In-house panels
- Running an online survey
- The online quality debate
- Summary of online quantitative research

Note, there are other forms of quantitative research that are not survey research, such as web analytics and scanner data. These are not covered in this part of the book.

Overview of Online Quantitative Research

The largest impact of online market research so far has been in the area of quantitative survey research, both in terms of volume and value. By 2008, according to the 2009 ESOMAR Global Market Research Report, the value of online quantitative market research was 20% of the global research total, which put it at about US$ 6.5 billion. By contrast, telephone quantitative accounted for 18%, face-to-face surveys for 12%, and qualitative research was reported as representing about 14% of global market research revenues.

It is important to note that the majority of quantitative research conducted via online data collection is not inherently different – in terms of the problems it is seeking to solve, the questions it asks, and the analysis that is conducted – from quantitative research conducted via other modalities, such as face-to-face or telephone.

This chapter is a high-level review of the key issues that confront market researchers and market research buyers when conducting online quantitative studies.

The topics covered in this chapter are:

- *Online survey process*
- *Surveys and the internet modality*
- *Finding respondents to interview*
- *Moving surveys online*

ONLINE SURVEY PROCESS

The typical process for the data collection phase for a project using online data collection is as shown in Table 1.1.

It is important to note that the timeline for an online survey is different from that of a telephone study, and even more so from that of a face-to-face study. In non-trivial cases, the total time required for an online survey will be shorter than that for a telephone or face-to-face survey, but this is not true of each individual stage. Compared with a face-to-face or telephone survey, Steps 2 to 4 of

Table 1.1

1	Create and agree with the client a draft questionnaire, often referred to as a paper questionnaire or a Word questionnaire (as in Microsoft Word).
2	Script the survey, i.e. create an online questionnaire, typically using one of the data collection systems.
3	Host the survey, i.e. install the questionnaire on a server attached to the internet and link it to a database to store the results.
4	Test and approve the survey.
5	Invite people to take the survey, for example send invitations to people on a database or to members of a panel.
6	Collect the data and monitor the progress of the survey, possibly tweaking elements such as quota controls.
7	Close the survey, download the data, remove and/or archive the online survey and data.

Table 1.1 tend to be slower for online projects, whereas Steps 5 to 7 tend to be faster (note, in a face-to-face survey Steps 2–4 would be the printing and dispatch of the questionnaires and Steps 5–7 would be the fieldwork and the data punching). Most of the time savings that are created by online surveys occur in Step 6.

In a face-to-face or telephone survey, the appearance of the questionnaire, which is dealt with in Step 2, is normally only a minor issue. Because the online modality is a self-completion mode, the layout, wording, and appearance of the survey are extremely important, as they are for postal surveys. Better designed, more carefully worded, more engaging surveys produce better data, but they also take longer to produce.

Step 3 in Table 1.1 is normally a quick process for online surveys and is analogous to the hosting of a CATI questionnaire for a telephone study.

Step 4 does not have a direct analogy in most face-to-face surveys; however, it can delay the online survey process. In all modalities it is a good idea to pilot a survey, but in online surveys the software and the layout have to be tested too, if major problems are to be avoided.

SURVEYS AND THE INTERNET MODALITY

The language of questionnaires and surveys tends to be derived from face-to-face research, with its references to questionnaires, pages, interviewers etc. However, each specific modality has its own features and characteristics that need recognising and utilising. This section reviews what we mean by different types of internet-related surveys.

WEB SURVEYS

In most cases, an internet survey is a questionnaire accessed via an internet browser (such as Microsoft's Internet Explorer or Mozilla's Firefox). This type of survey is commonly known as a web

survey and is typically created and operated via specialist online survey software (such as the products from companies like Confirmit, Nebu, or Voxco). The survey can be provided via a range of tools, for example HTML, JavaScript, and Flash, and can appear as either one long page or as a number of screens, with each screen showing one or more questions.

Later chapters of this book cover the choice of web interviewing systems, the design of web surveys, and managing web survey projects.

There are alternatives to the web survey, but none of the alternatives is in widespread use. Examples of the alternatives include email surveys and downloadable surveys, which are covered in the sections below.

EMAIL SURVEYS

The term 'email survey' normally refers to a survey that is emailed to respondents. An email survey does not normally refer to studies where the survey is hosted on the web and an invitation is emailed to the respondent; this is normally referred to as a 'web survey'. However, email surveys have become much less common than they used to be and the term 'email survey' may have become more ambiguous.

In the very early days of online research, email surveys were popular because they did not require people to be online whilst they completed the survey. The internet at that stage tended to be very slow and most users were paying for access to it by the minute. Also, in those early days there were many people who had email access but who did not have access to the web, particularly from PCs they accessed whilst at work.

There are two ways to conduct an email survey. The first is to include the survey in the body of the email and the second is to send the survey as an attachment to the email.

Email surveys in the body of the email

Email surveys in the body of the email further divide into two main varieties, text emails and HTML emails.

A text email sets out the questionnaire as part of the email in the same way that a message is typed into an email. The respondent clicks 'Reply' and then scrolls down the email, entering their responses. The main drawbacks with this approach were that the survey tended to be boring, restricted, and respondents could miss questions, answer inappropriately (for example picking two options where only one should be selected), or even delete parts of the questionnaire. Interview software exists to draft these surveys, to conduct the mailing, and to interpret the replies.

An HTML email survey uses the power of HTML to create a more interesting survey. An HTML survey can look appealing. For example, it can use standard features such as radio buttons and check boxes, and can include a SUBMIT button at the foot of the email, making the survey more intuitive.

The main problem with HTML surveys is that many people's email filters will either prevent HTML emails from getting through, or they will convert them to plain text.

One area where email surveys have a specific benefit, compared with web surveys, is within a large organisation, for example when conducting a staff survey. In a large organisation a web-based survey can cause many recipients to log into the survey at the same time, potentially causing bandwidth problems. An email study distributed within a company will, in many cases, spread the load more evenly.

One special case of an email survey is where an email service is utilised to poll people's views, for example the voting buttons in Microsoft's Outlook can be used to gather people's views.

Email surveys as an attachment

In the early days of the internet it was quite acceptable to email a survey as a piece of executable code. These emails arrived on the respondent's machine, the attachment was opened, the survey completed, and it then emailed itself back to the project sponsor. The general reluctance to accept executable code via email has resulted in this form of survey becoming rare.

DOWNLOADABLE SURVEYS

A downloadable survey is one that is downloaded from the internet to a local device and the results are then sent back to the server at the end of the data collection process. Downloadable surveys tend to be implemented for mobile devices such as smartphones rather than for PC-based surveys.

One more recent innovation is to include a mobile, downloadable survey as part of a wider project, such as e-ethnography (a subject of a later chapter), but this remains relatively rare.

FINDING RESPONDENTS TO INTERVIEW

Initially, one of the main reasons that online data collection was held back was because there was no reliable and scalable method of contacting potential respondents. Face-to-face research was able to draw on sources such as electoral registers, postal address files and similar sources. Telephone was able to draw on directories and RDD (random digit dialling), but there was not (and still is not) a directory of who is on the internet.

However, through the 1990s and beyond, methods have been created that provide ways of finding respondents for online surveys. It should be noted that most of these innovations have required some changes in the assumptions about how sampling is conducted, a topic that is discussed later.

The key methods of contacting respondents are:

1. Online access panels
2. Client databases
3. Marketing databases
4. Client panels
5. Website visitors
6. River sampling

The following sections briefly define and outline these six approaches. The major techniques, and the implications of working with them, are dealt with in greater detail in later chapters. This section then finishes with a brief review of additional methods that have been used to find samples.

ONLINE ACCESS PANELS

The two big changes that facilitated the wide scale adoption of online data collection were the development of online access panels and the willingness of a large proportion of market research buyers to move away from the assumption of random probability sampling. This first of these two changes was obvious to both suppliers and buyers of research. It is less clear that the loss of the ability to claim random probability sampling was as widely understood.

Online access panels are also known as access panels, online panels, or simply just panels.

International online access panels, such as SSI, eRewards, LightSpeed, and GMI, are able to provide online samples for most of the developed markets (developed in terms of markets with a large volume of market research). In addition, in most developed markets there are a variety of local panels.

There are a number of different ways of working with online access panels, but the two most typical ways are:

1. The researcher provides the panel company with a sample specification and a link to a survey that has been scripted and hosted. The panel company then invites its members to take part in the survey.
2. The researcher specifies the survey and the sample specification and the panel company scripts the survey, hosts it, and provides the sample.

Because of the importance of online access panels there is a chapter specifically on working with them later in the book.

CLIENT DATABASES

Most organisations have some sort of customer database. These tend to vary from a very sophisticated CRM database of all customers (for example for an online retailer) through to a basic list of email addresses supplied by customers on an *ad hoc* basis.

The typical method of conducting research with a client database is to send a survey invitation to a sample of members drawn from the database. The most common way of sending the invitation is via email but there are other ways; usually the invitation will include a link to a web survey. There is more about working with client databases later in the book.

MARKETING DATABASES

There are a large number of organisations who hold databases for marketing purposes. The majority of the customers for these databases are direct mail companies, but they are sometimes used to provide market research samples.

The main concerns that market researchers have expressed about this approach are that: (1) there are few constraints about who is on these lists; (2) none of the quality guidelines which have been agreed by the research industry (see quality notes later) apply to these databases; and (3) the people on marketing lists may confuse marketing and market research, or competitions and market research.

Marketing database companies do not usually offer a scripting and hosting service for surveys, so the typical way of using them is that the database company sends out an email invitation to the survey, using an invite agreed with the researcher, including a link to the web survey.

Researchers who have used marketing databases have reported that they have experienced problems with duplicate responses (the same person answering the survey more than once, presumably because they are on the database more than once), and of a higher than expected number of people interested in prizes and competitions.

The key differences between an access panel and a marketing database are that:

(a) the access panel is only used for market research, it is not used for marketing

(b) members of an access panel know they are on the panel and that they will only be contacted for market research purposes

(c) most market research online access panels have signed up to guidelines, such as the ESOMAR guidelines

When using marketing databases, the researcher should make an effort to ensure that the people on the lists have given the appropriate permissions and that appropriate efforts have been taken to

screen out children (or to obtain the relevant parental permission). The response rates from these lists tend to be lower than those from online access panels. Marketing databases have different pricing policies, with some charging by invitation, some by survey starts, and others charging only for completes.

Despite criticisms of marketing databases, they have been used in the past to help create online access panels.

CLIENT PANELS

A number of clients have created their own research panels. The key differences between a client panel and a client database are the same as the differences between an access panel and a marketing database.

Client panels vary in size from a few thousand members to tens of thousands of members. Some panels are managed in-house whilst others are managed by third-party companies. There is more about working with client panels later in the book.

WEBSITE VISITORS

The people who visit a website can be sampled for quantitative samples. For example, visitors to an online news service can be asked to do surveys about the website, the brand providing the news service, or the news service. This topic is covered in greater depth in Chapter 15.

There are two ways of soliciting the respondents: (1) using popups (or something similar such as overlays or pop-unders); or (2) by placing invites on the website.

Popups surveys can also be used to recruit respondents for general surveys, but this approach is better categorised as being a river sampling technique, especially if multiple sites are used.

RIVER SAMPLING

The concept behind river sampling (sometimes called real-time sampling) is to find and recruit respondents as and when they are needed, rather than keep going back to the same small sub-set of people who have chosen to join a panel or community.

Some of the proponents of river sampling claim that it more closely approximates to random sampling, compared with online access panels. However, most of the advocates concentrate on the 'freshness' of river samples, rather than any truer sense of representativeness. They point out that members of online access panels may complete upwards of 100 surveys a year. By contrast, fresh respondents,

recruited by river sampling, are less likely to have had their responses affected by having completed large numbers of surveys.

The proposition that 'fresh is best' is one that has some face validity but the research conducted by the ARF's 'Foundations of Quality' project suggests that the number of surveys that somebody completes is not a problem. So a preference for a 'fresh' sample has to be considered a personal preference at the moment, rather than an evidence-based decision.

Examples of river sampling include:

- *banner ads*
- *popups*
- *interstitials*
- *overlays*
- *interaction devices within social networks*

River sampling represents a minority of all online recruitment, although some companies such as DMS Research in the USA have made a major feature of using it. River sampling is at its strongest when it can draw a sample (i.e. intercept people) across a wide range of sites, rather than concentrating on just one or two.

OTHER SOURCES OF SAMPLE

In addition to the methods outlined above, a wider range of alternatives have been trialled:

- *telephone recruiting*
- *SMS recruiting*
- *outdoor adverts, e.g. billboards*
- *ads in newspapers and magazines*
- *postal invites*
- *URLs posted on products or on receipts*
- *public access points, e.g. cyber-cafés or kiosks*

However, none of these has proved generally useful, even if they have proved useful in specific cases. Generally these methods produce low levels of response.

MOVING SURVEYS ONLINE

Moving a survey online describes the process of taking an existing study, for example one using face-to-face, telephone, or post, and moving the data collection to the internet. People have been moving surveys online since the late 1990s, so there is a substantial amount of learning that can be drawn on.

One of the implications of moving an existing survey online is the need to deal with legacy issues. Two common examples of studies with legacy issues are:

1. **Tracking studies**. For example, looking at elements such as brand, advertising, or customer satisfaction over time. Changing the data collection modality might result in a loss of ability to track changes.
2. **Studies that use norms or benchmarks to help interpret findings**. For example, concept testing, advertising pre-tests, and opinion polling often use weights or procedures based on historical projects. When a study moves online the weights or procedures might need to change.

Advice

The key questions that need to be addressed when moving a survey online are:

- *Can the right sample be contacted online?*
- *Is the online sample large enough for the project?*
- *Is the online sample different from the existing sample?*
- *Are all the current questions capable of being asked online?*
- *Would the current questions generate different answers online?*
- *Should the new survey minimise the differences from the old study, or maximise potential improvements?*
- *Can the new study be piloted or run in parallel with the old study?*
- *Can any differences in the results be modelled?*

Taking all the points outlined above, it can be seen that when migrating a project from one modality to another there are three principal causes of changes to consider:

1. **Population effects**. The population that is available to an online research survey might be different from the population available to an offline survey.

2. **Sample effects**. The sample of people who are likely to be attracted online might differ from the sample likely to be attracted offline.

3. **Method effects**. The same person might respond in a different way to an online survey than to the way they would have responded to an offline survey.

The following sections address the points described above. When looking at potential differences the researcher should keep in mind that different does not mean worse. Different can be better, it can be worse, or it can just mean different.

CAN THE RIGHT SAMPLE BE CONTACTED ONLINE?

One key issue that needs to be addressed is whether the people who have been sampled in the past tend to be online and whether the proposed survey is likely to find them if they are. To answer this question the researcher may need to consider whether the sample is going to be sourced from an internal database or from an online access panel.

If the sample for the online survey is different from the previous sample (e.g. younger, richer, more likely to buy online, watching less TV), then it is likely that the results will be affected by the change to online.

One example of a case where online is not an appropriate modality is where the research objective is to estimate the usage of the internet. To find out internet penetration figures, another modality must be used. Even the scale of usage is hard to measure via online surveys, because the people who are reached will tend to be disproportionately and systematically heavier users of the internet.

Another example of a study that tends not to work online is a study looking at how people book flights from travel agents. The people online are likely to be much more likely to book flights directly and to potentially only use travel agents for complex journeys, whereas people without internet access may use travel agents for a wider range of services.

IS THE ONLINE SAMPLE LARGE ENOUGH FOR THE PROJECT?

If the study is a large tracking study, for example a weekly brand tracking study, the sample source needs to be large enough to be able to conduct the survey without re-sampling the same respondents too frequently and without having to shift to an alternative source. There are a wide number of studies showing that changing from one source to another, for example from one panel to another, is likely to produce changes in the results unrelated to changes in the population.

IS THE ONLINE SAMPLE DIFFERENT FROM THE EXISTING SAMPLE?

Even if the people who can be contacted online appear to be similar to those who were contacted previously, the people who agree to complete the survey might be different. The online sample, even if matched in terms of demographics, may show differences in terms of attitudes, beliefs, or experiences.

ARE ALL THE CURRENT QUESTIONS CAPABLE OF BEING ASKED ONLINE?

Some questions that are easy to ask in a face-to-face interview or over the telephone can be problematic to ask online. A very good example of this is the common research practice of asking for an unprompted list, followed by a prompted list, for example, an unprompted list of brands followed by the prompted list of brands. When an interviewer is present the unprompted list can be supported with probing. For example, if the question is something like '*Which soft drinks do you drink?*', a respondent might say 'Coke', at which point the interviewer can probe by asking, '*What type of Coke is that?*'

When conducting an online survey, an unprompted list has to be asked as a set of open-ends. This has several implications for the process. The first is that people do not type in names as accurately as a computer interprets them. If somebody types 'Coka-Cola' the computer won't necessarily recognise it as Coca-Cola, and the survey is very unlikely to probe for which variety of Coke. The problems are then compounded at the prompted stage. In a face-to-face or telephone interview the interviewer normally fills in the items that have already been spontaneously mentioned, and then prompts for the rest. In an online survey the respondent typically types in open-ended responses in the unprompted section and then has to answer the whole list, selecting again items mentioned in the unprompted question.

A researcher converting a survey online needs to review the current questionnaire to see whether changes are needed to accommodate the restrictions of the online medium.

DO THE CURRENT QUESTIONS GENERATE DIFFERENT ANSWERS?

Even when a question can be asked in the same way online, the answers may not be the same. It has been suggested that one of the differences between interviewer mediated surveys and self-completion is that in self-completion the respondent is more honest, which is often manifested as lower scores on questions such as likelihood to buy (Comley, 2002). It has been suggested that the absence of an interviewer results in respondents being less tempted to provide socially acceptable answers.

In some cases there may be data available to indicate whether a specific question and wording result in different answers online. However, in most cases it is necessary to conduct some sort of pilot or comparison.

SHOULD THE NEW SURVEY MINIMISE THE DIFFERENCES FROM THE OLD STUDY, OR MAXIMISE POTENTIAL IMPROVEMENTS?

Moving a survey to the internet often provides an opportunity to improve the results, for example by using better stimuli than is possible via telephone, by using better randomisation and complexity than face-to-face, or because of the increased honesty that results from not having an interviewer present.

However, better results are different results, which can be a problem. For example, if staff bonuses are linked to the results of a satisfaction study, changing the results (even if the new results are more reliable or valid) may have significant repercussions for the organisation.

One option that is often adopted when moving a survey online, is to accept that there will be a break in continuity and use the change to conduct a major review of the survey, removing redundant and less effective questions and possibly adding new questions.

CAN THE NEW STUDY BE PILOTED OR RUN IN PARALLEL TO THE OLD STUDY?

The best practice for moving a survey online is to pilot the new online survey in parallel with the previous study. This process can identify any problems with the online implementation and allow any differences in the results (including issues like response rates and respondent satisfaction) to be assessed. However, there is often a significant cost implication and it can delay the change.

CAN ANY DIFFERENCES IN THE RESULTS BE MODELLED?

If the new online survey is run in parallel with the previous study for a period of time, it may be possible to model the differences between the two studies. For example, if the online sample is 10% more satisfied than the offline sample, the old data can be modelled (by adding 10% to their satisfaction scores). Note, best practice is to model the old data, not the new data. Modelling the old data only requires one set of modelling, but if the new data is modelled (to make it like the old data) then the modelling has to be run on all new waves of the data.

HIGHLIGHTING CHANGES IN THE PARADIGM, CONFIDENCE, AND IMPLICATIONS

When the data collection modality for a project changes there may be obvious differences, for example differences in key scores. However, there may also be less obvious changes, for example changes in the underlying paradigm, or the level of confidence that users of the research can have in the findings, or the managerial implications of these changes for the client.

If the study to be migrated to the internet is currently based on the paradigm of random probability sampling, this may change when the study goes online (unless it is a customer database study and all/most of the customers are online). If the existing study is based on an assumption that a cross-section of the population is interviewed and that quotas are used to control the nature of that cross-section, then the online study may fit the same paradigm.

When moving a survey online it is a good idea to look at the claims that have been made about the reliability and validity of the study. It may well be that the current claims need modifying, irrespective of whether the survey stays offline or moves online. One change that might be helpful is to change claims about sampling error away from validity (i.e. actually represents the population) and towards reliability (i.e. if we asked this sort of sample the same questions again, this is how likely we are to get the same results).