

## DEVELOPING NEW QUALITY INDICATORS IN SOCIAL SURVEYS

Lucy Haselden and Amanda White<sup>1</sup>

### ABSTRACT

Rather than rely on traditional measures of survey quality such as response rates, Social Survey Division of the UK Office for National Statistics, has been looking for alternative ways to report this. In order to achieve this we have mapped out all the processes involved throughout the lifetime of a survey, from sampling and questionnaire design through to producing a finished report. Having done this, we have been able to find quality indicators for many of these processes. By using this approach, we hope to be able to appraise any changes to our processes as well as informing our customers of the quality of the work we carry out.

KEY WORDS: Quality indicators; process mapping; surveys

### 1. INTRODUCTION

The quality of a survey is a matter of great interest to both survey organisations and those that commission surveys. Commissioning organisations have always needed to know whether the data they are using is fit for purpose and whether the benefits offered by suggested improvements to survey methodology will warrant the costs they will have to pay. However, in an age where more and more surveys are open to competitive tender, commissioning organisations are also having to weigh up the quality of data offered by different bids and so there is increasingly a need for them to have a sound means of making these comparisons.

Survey organisations also need to know about the quality of the data they are producing. Survey organisations often have to bid for the work that they carry out and an organisation that is able to make strong, verifiable assertions about the quality of its data will be in a better position for individual bids than one that is unable to produce evidence to back up its claims.

However, the advantages of measuring quality go beyond marketing. Having objective ways of measuring quality enables comparisons to be made within the organisation, both between surveys and before and after innovations. This will make it easier to ensure that best practice is adopted throughout the organisation.

Social Survey Division (SSD) in the United Kingdom's Office for National Statistics (ONS) have therefore been trying to find structured ways of measuring the quality of their data that can be used across different surveys and can be both presented to clients and used internally as a management tool for improving the quality and efficiency of our procedures.

This work is still very much in progress but the paper sets out to describe the methodology we have used to find these quality indicators, as well as setting out some of the quality indicators we plan to use and how we are going about implementing them.

### 2. TRADITIONAL MEASURES OF QUALITY

Traditionally, approaches to survey quality focused on minimising survey error and looking at how closely the survey estimate conforms to the true value.

One way of measuring this was through concepts such as Mean Square Error (Groves 2001) Mean Square Error (MSE) may be defined as  $E = (y - X)^2$  where  $y$  is the estimate of  $X$ , that is MSE is the square of the difference between the true figure and the estimate.

There are several problems with this approach. The first is that it relies on knowing the true value behind the estimate. Clearly, we will rarely know this but there are occasions where, for certain variables we may have access to better estimates (e.g. estimates from the Census or other surveys with very large sample

---

<sup>1</sup> Lucy Haselden and Amanda White, Office for National Statistics, 1 Drummond Gate, London, UK SW1V 2QQ,

sizes and high response rates). However, these “better” estimates will never be perfect and will only be available for certain variables which will often not be the ones we were principally setting out to measure.

The second problem is that it simply measures how closely the survey estimate conformed to a better estimate but does not take into account any aspects of design such as whether the survey was asking the correct questions in order to answer its research objectives.

Furthermore, any estimate of MSE can only be made after the survey has been completed and the results are available. This means it cannot be used as a basis for ensuring the quality of the survey during its life time so that problems can be rectified as they happen. Indeed, it gives no indication of where any error occurred (e.g. in sampling, editing, coding) but merely that there was error. As such it is not a very good tool for assisting survey managers in implementing best practice.

The impracticality of calculating Mean Square Error and its limited usefulness means that survey organisations have used other measures of quality. The most common measure used is response rates. These are very useful because they give some indication of the representitiveness of the survey sample and the amount of bias we are likely to expect. They are also relatively easy to calculate and can be monitored whilst the survey is in the field, with final details of response rates being handed over with the data.

However, response rates only measure one aspect of data quality: they do not tell us anything about how good the sample design is, whether the sampling frame was complete, whether the questions asked fulfilled the research needs, whether the data was edited properly or any of a number of other things that might effect the quality of the final data set.

In addition, response rates are measured very differently across different survey organisations and there are often differences even within organisations. A separate but related project which we are carrying out in conjunction with the National Centre for Social Research aims to look at ways of harmonising the calculation response (Lynn et al 2001).

Neither response rates nor MSE seemed to be adequate measures of data quality that could be used by commissioning organisations, to assess overall value for money and fitness for purpose, or by survey managers as a tool for implementing best practice. So we needed to find another measure or package of measures which would come closer to fulfilling these requirements.

### **3. THE PROCESS APPROACH**

One way of looking at the quality of survey data that has been mentioned in various pieces of literature, is to look at the “process quality”. We can view the creation of any product as a process with inputs and outputs. For example, we can view making a cake as a process with inputs (the ingredients, the oven, the utensils and the cook’s time) and an output (the cake). However, we can break that process down into a number of sub-processes which each have their own inputs and outputs. Using the cake analogy, mixing the ingredients might be considered one sub-process which has inputs (the raw ingredients, utensils etc) and an output (the cake batter). The output from that process will then be fed forward to become the input for the next process so that, in this example, the cake batter becomes the input for the baking process.

In fact it is often the case that there is more than one input for each of these processes and there is often more than one output. Nevertheless, each sub-process has to produce a quality product for its customer further down the chain. The quality of each of these sub-processes will affect the quality of the overall process. A process approach to quality will therefore look at the quality of the outputs of each of the sub-processes. If the quality of all these outputs is high, and the processes are the correct ones, then the quality of the final output should be high. Likewise where the quality of the output of any of the sub-processes has been low, there is likely to be a corresponding drop in the quality of the final product.

This process approach has a number of advantages. Firstly, it allows the survey manager to be responsive to problems that occur during the survey. If a problem has occurred during a particular process, early identification of that problem may mean the survey manager can take measures to counter those problems and still produce high quality data. Moreover, if that process is commonly used by other surveys, they will be able to fix any problems before other surveys experience them.

A process approach will also allow the survey manager to evaluate changes to certain survey processes. For example, if there is a change in the editing procedure, the survey manager can look and see what

advantages or disadvantages that change in survey procedure brings. We therefore decided to try to find indicators which would assess the quality of all the major processes associated with running a survey.

## 4. MAPPING SURVEY PROCESSES

### 4.1 Method

In order to find quality indicators for each of our processes we needed to know:

- What each of the processes were, their inputs and outputs and how they fitted together;
- Who the process owners were as well as their customers;
- What constituted quality for each of the processes.

The first stage was to map the survey processes. In order to do this we needed to find out what all the processes were and how they fitted together. This was quite a complex task because different surveys use slightly different methodologies and we wanted a way of describing what we do in generic terms wherever possible. In order to do this we used a number of sources of information.

The first source of information was the documentation that already existed within ONS. A survey procedures manual gave a broad account of how to run a survey within ONS and in addition, each survey had its own documentation that related specifically to that survey. This proved a useful starting point but did not provide all the information we required: much of the documentation was not written from a process perspective so it was not always clear what the inputs and outputs were, nor did it always identify the process owners and customers. Survey managers were also able to provide details about the processes on their surveys but although they were able to provide good information about the overall workings of their survey, the information that they provided was very much from a research point of view. We therefore, decided to supplement the information obtained from the documentation and survey managers, with first hand experience from those people who actually carried out the tasks or process owners. This would ensure the process map was as up to date as possible and accurately reflected the reality of how people carried out their jobs.

As we interviewed the survey managers and process owners, we drew up a process map including all the major tasks commonly carried out on surveys.

## 5. OUR PROCESSES

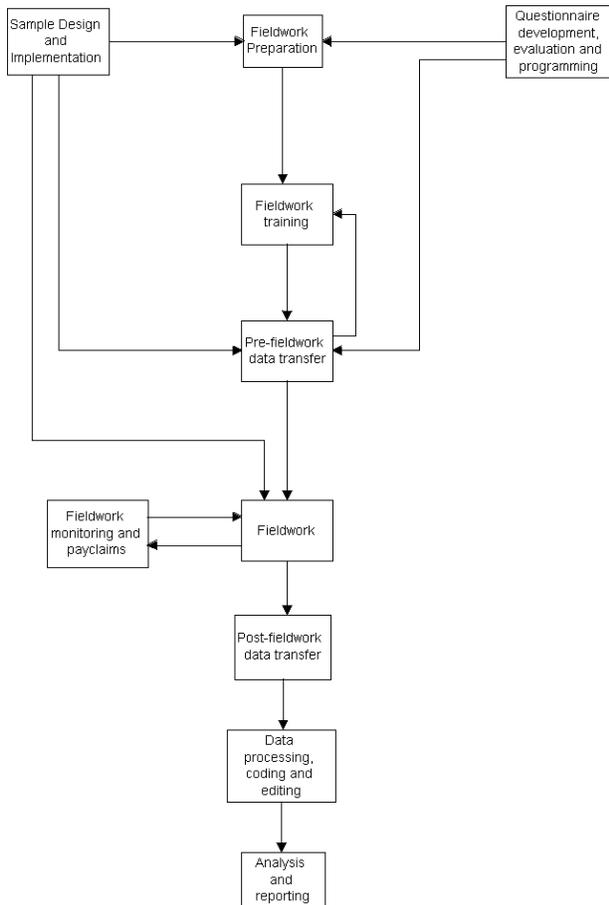
When we had drawn up the map, it became clear that we had over 100 different processes taking place on any survey. Some of these happened sequentially, some simultaneously and most of them had more than one input and/or more than one output. We therefore had to find a way of presenting these processes in a meaningful way. One way of achieving this was to define ten main processes, each having their own inputs and outputs. All the other processes could then be considered sub-processes of these. The main processes were:

- **Sample design and implementation** – this includes all the sub-processes concerned with designing and drawing the sample and is discussed in more detail below.
- **Questionnaire development, evaluation and programming** – this includes designing the questionnaire, carrying out any question testing and programming the Computer Assisted Interviewing (CAI) instrument or designing the paper questionnaire.
- **Fieldwork preparation** – this covers selecting interviewers and ensuring that they have all the materials they need in order to be able to carry out the interview.
- **Fieldwork training** – this includes any briefings that interviewers receive as well as any other training that they take part in (such as reading interviewer instructions).
- **Pre-fieldwork data transfer** – this includes sending all the details of the sample and the questionnaire to the interviewers.
- **Fieldwork** – this covers all the sub-processes involved with actually carrying out the interviews.
- **Post-fieldwork data transfer** – this describes how all the data that the interviewers have collected are sent back to our head office.
- **Fieldwork monitoring and pay claims** – this involves all our processes for monitoring our interviewers and finding out how much time they have spent in the field.

- **Data processing, coding, editing and checking** - this covers a series of sub-processes in which the data are cleaned, coded and checked in order to produce a final data set which is ready for analysis.
- **Analysis and reporting** – this involves running tables and frequencies as well as any modelling and writing up the results into a final report or presentation.

A map was drawn up showing how each of these main processes fitted together (see figure 1). In addition, additional maps were drawn up showing how each of the processes listed above were carried out and what sub-processes were involved

**Figure 1 Survey process map**



Outputs from ‘Sample design and implementation’ become inputs to ‘fieldwork preparation’, ‘pre-fieldwork data transfer’ as well as directly into fieldwork. It is possible to see from the diagram how all the other main processes feed into one another.

## 6. IDENTIFYING QUALITY INDICATORS

Once we had drawn all the process maps and knew about all the processes that took place, we could set about looking for quality indicators. We wanted to find quality indicators for each of the major processes. At this stage, we wanted to include any possible quality indicators whether their primary use would be as a management tool or as a quality indicator to be passed on to clients, or both.

We set about this by asking the process owners what they needed in order to be able to do their job well, using the process maps as a guide. It became clear, that the quality of the output from one process often did not become apparent until several processes down the line. For example, the immediate customer for the programmed questionnaire is ‘pre-fieldwork data transfer’, that is those people who are responsible for data transmission, as it is at this stage that the questionnaire is distributed to the interviewers and put on their laptops. The process owner for ‘pre-fieldwork data transfer’ will know about some aspects of the quality of

the questionnaire such as the timeliness of its delivery and whether they are able to transmit it to interviewers. However, they will not be aware of some aspects of the questionnaire quality that will be felt by less direct customers of this process. For example, the interviewers will need the questionnaire to be well designed and to flow well in order to be able to do their job.

By asking process owners what they need to be able to do their job well, we can determine what all the aspects of quality are for each process, even if they are not apparent to the direct customer for that process.

The following sections list the main processes that occur in a survey together with some of the ways in which they effect the overall quality of the survey and how these can be measured.

## **6.1 Sample design and implementation**

### *Precision*

The sample design must take into account the level of precision required by the users of the data. The precision of the data will be determined by the size of the achieved sample together with other design aspects. If the sample is stratified, using appropriate stratifiers, this should increase the precision of the sample around key estimates. Conversely, if a sample is clustered, this will reduce the precision of the sample. The effect of stratification and clustering on the precision of a particular estimate is described by the design factor or deft.

One way of assessing the precision afforded by the sample design is to examine the defts for key survey estimates to see whether these match with expectations. If we calculate the defts for particular estimates and know what the achieved sample size was, we will also be able to calculate the effective sample size for that estimate. The effective sample size is the size of the simple random sample that would be required to achieve the same level of precision as the sample with stratifiers and clusters. These two measures (design factors and effective sample size) can then be used as quality indicators for the sample design which can be passed on to customers.

### *Cost of fieldwork*

The sample design should also take into account the cost of fieldwork. By clustering a sample in a certain geographical area it is possible to reduce field costs. The aim will always be to achieve the required level of accuracy at the lowest cost. We can monitor whether this is being achieved by looking at the cost per unit effective sample size for key surveys estimates. This can be a useful management tool for determining the effectiveness of different sampling strategies.

### *Clarity of sample instructions*

Once the sample has been drawn, details need to be sent to interviewers. Normally, in SSD, this will include details of the addresses at which they should interview as well as how to carry out any sampling in the field. The details of the addresses should be clear and unambiguous as should any sampling instructions.

We can measure the quality of the sample details that are provided to interviewers by monitoring the number and types of queries that we are receiving from interviewers. Interviewers are asked to telephone a helpline number if they have any queries about the sample details. By logging all their calls and producing a report of the number of calls for each survey of a particular type, we are able to check whether interviewers are receiving adequate sample instructions as well ensuring that they are well briefed about how to interpret and follow the instructions they receive.

## **6.2 Questionnaire development, evaluation and programming**

### *Ease of understanding and answering questions*

Interviewers, and, more importantly, respondents should be able to understand the questions they are asked as well as understanding how they should answer. We can help ensure this happens by making sure that the questions are adequately tested. The Question Testing Unit in ONS provide a report for any work that they carry out, and whilst this is not a hard figure, the report can be a useful quality indicator for clients and can be included with any other quality information they are given.

Another way of looking at the quality of questionnaire design is to monitor the number of “don’t knows” and “refusals”. A high number of don’t knows or refusals for a particular question might indicate that there is a problem with the question: perhaps respondents don’t understand the question or are unsure about how to answer it or perhaps it is asked in an insensitive way. It may also happen when the question asks for

information that respondents simply do not have. In all these cases, the problem might be alleviated if the question is better designed.

#### *Accuracy of CAI programming*

Most surveys carried out by SSD use CAI. CAI has many advantages over paper and pencil as it ensures that interviewers all follow the same routing and allows some editing to take place in the field so that if respondents give unlikely or improbable responses, the correct answer can be established with the respondent rather than in the office by someone who cannot know all the respondent's circumstances. Furthermore, edits that take place in the office can be time consuming and so there may be some advantages in speed of data delivery in editing as much of the data as possible whilst in the field.

However, the gains in quality that can be achieved by using CAI depend upon how well the CAI instrument has been programmed. We can assess the accuracy of the CAI programming by recording the amount of office based editing that was required. Although there are occasionally some edits which are best carried out in the office, some office based editing may be required because either they were omitted from the CAI instrument or because there was an error in the CAI program that resulted in respondents being misrouted. In either case, a substantial or increased amount of editing might indicate that improvements can be made to the CAI program.

If there are any problems with the programming of the CAI instrument, interviewers are usually quick to spot and report them. Once a problem has been reported, we aim to correct and redistribute the program as soon as possible. Although such redistributions are rare, we want to monitor how often they happen and why.

### **6.3 Fieldwork preparation**

#### *Fieldwork allocated efficiently*

In order to minimise costs, quotas of work should generally be allocated to the closest available interviewers and those interviewers should cover the whole country. If interviewers are routinely having to travel long distances to quota areas, this suggests that either work is not being efficiently allocated or insufficient interviewers are being recruited in some areas. We can therefore monitor the efficiency of the allocations by recording the number of miles that interviewers travel to the area. This information can be a useful tool for project managers and those that recruit interviewers.

#### *Interviewer experience*

Interviewers are sometimes allocated to work according to their experience, especially where the survey has a complicated design or where response has been falling. It is generally thought that experienced interviewers can improve response rates and will be better able to deal with complex survey designs. We can measure the experience of our interviewers in terms of the number of years they have worked for us and the number of quotas, both on a particular survey and in general, that they have worked on. This information might be of interest not only to project managers but also to clients who want to see that their data has been collected by experienced interviewers. It will also allow us to examine the relationship between interviewer experience and response.

### **6.4 Fieldwork training**

#### *Quality of briefings*

Before interviewers begin work on a survey for the first time, they are briefed about that survey. It is important that interviewers attend the briefings and they get as much out of them as possible. In order to gauge how effective briefings have been, we can ask interviewers to assess the quality of them by filling in a brief questionnaire. The answers to the questionnaire may be used as one way of assessing the quality of the briefing. Another way of assessing how well briefed interviewers are is by testing them on the most important aspects of the survey. High scores for the test suggest that the interviewers know their material well and have been well briefed whereas low scores suggest there are areas which need to be covered in greater depth during the briefing.

### **6.5 Pre-fieldwork data transmission**

#### *Questionnaire and sample data despatched intact and on time*

SSD's survey computing branch uses computer links to send interviewers the questionnaire and sample details. Interviewers want to receive this information on time and intact. On occasion the data files sent to

interviewers become corrupted or lost. This will always cause some problems and may even delay the start of fieldwork. We therefore want to record how often this happens and the cause of the problem.

## **6.6 Fieldwork**

### *Response rates*

Ideally we want to interview at every address we have sampled in order to remove as much bias as possible. We cannot always achieve this because some addresses are ineligible to take part in the survey, sometimes we cannot contact anyone at the address, and sometimes the occupants refuse to take part. Measuring the response rate will give some indication of how good the fieldwork was and how likely we are to have bias. The ease of collecting information about response rates and the information they convey about likely bias means that clients are usually very interested in response rates and often set final response rate targets for surveys. Survey managers also find it useful to monitor response rates whilst the survey is in the field in order to ensure that interviewers are making sufficient progress with their work and to check whether any remedial action needs to be taken in order to meet targets.

We also record and publish details of the ineligibility rate. Although there might be slight variations to the ineligibility rates across the country, we would generally expect fairly constant eligibility rates over time and between surveys, unless there are differences in the definitions of what constitutes “eligible”. Where there are substantial unexpected differences in eligibility rates, this might suggest there is a difference in the way interviewers are recording it.

### *Interview length*

Interview length contributes both to the cost of the survey as well as to respondent burden and hence to the likelihood of respondents taking part. For both of these reasons, it is helpful for survey managers to monitor the interview length. Furthermore, recording interview length is a way of checking interviewers’ work. We might want to investigate cases where the interview took place in an extremely short time as this may imply bad interviewing techniques or possibly, in the worst case, interviewer fraud. Where CAI is used, we can calculate the length of the interview from time stamps in the program. It is very difficult for interviewers to alter this and so this is an impartial and accurate way of recording interview length.

### *Quality of the interview*

Interviewers are asked to carry out interviews in particular ways: they must stick to the questions on the questionnaire and must not lead the respondent into answering in a particular way.

By carrying out probity checks, we can ensure, not only that the interviewer is interviewing at the correct households but also that they are following instructions given to them.

### *Efficiency of interviewer work patterns*

Interviewers are asked to record details of all the calls they make at each of the addresses that they are assigned. They record details of the time they visited the address, the outcome of each visit, as well as making any other relevant notes. We can examine these calls details to ensure that interviewers are working efficiently and are calling at addresses at several different times of day before recording them as a non-contact.

## **6.7 Post fieldwork data transfer**

### *Loss of data*

When interviewers send data back, we want to ensure that all the data they send is properly received. Occasionally, data are lost or corrupted when sent to the office. We can record how often this occurs to indicate the quality of this process.

## **6.8 Data processing, coding and editing**

### *Accurate data entry and coding*

Although we generally collect data using CAI, we do occasionally use paper questionnaires (e.g. for postal surveys, for self-completion sections of a questionnaire or for diaries). Data collected in this way must be entered into the computer in order to be analysed. Any errors at this stage will affect the accuracy and quality of the data. It is hard to find ways of measuring the accuracy of the data entry, we can record what methods were used to avoid incorrect data being entered (e.g. double keying) and we can carry out occasional exercises to compare the data entered with that on the original forms and count the number of discrepancies. We can then use these reports as a quality indicator.

Most questions are coded by the interviewers during the interview, although some coding work may be done outside the interview. In both cases the quality of the coding is important. We can carry out methodological work to assess how consistent the coding work is and whether it is done in line with instructions.

Such exercises examining the quality of data entry and coding are expensive to carry out and cannot be done routinely. However, every time such a study is carried out, we can quote the results as a quality indicator for this process.

## **6.9 Analysis and reporting**

### *Interesting and well written analysis*

Most customers want interesting, well written and presented analyses. It should not contain any errors or inconsistencies and be received on time by the customer. Although it is hard to get any objective measures of this, the results of customer satisfaction questionnaires can give us some indication of the quality.

## **6.10 Overall**

### *Timeliness*

Many customers cite timeliness as an important measure of the quality of the survey: customers want to receive the data and/or report in time for any work that they would like to carry out and before the data might become out of date. Similarly, the customers for all the different processes want to receive their inputs in a timely way. We aim to store all the key dates on the timetable and record whether these were met or not and if not, why not and how late is it. Timeliness of outputs from individual survey processes may be a useful management tool whereas the timeliness of the delivery of the final outputs might be seen as a quality indicator of the overall survey.

## **7. IMPLEMENTING QUALITY INDICATORS**

Once we had identified quality indicators for each process, we needed to implement them. The first stage of this was to find out where all the information was held and then to write a program which would gather this information and present it in a way which is meaningful and consistent across surveys. In this way, project managers would have access to reports for their survey and would be able to make comparisons between their performance in different time periods as well as being able to make comparisons between different surveys.

In order to ensure that the programs we are writing are as useful as possible to survey managers, we have tried to involve all the interested parties in writing the specifications, taking account of their requirements and concerns. One of the main concerns that has arisen is how these indicators should be interpreted. For example, a rise in the cost per unit effective sample size might arise from a change in the sample design but it could equally arise from a change in fieldwork practices. Similarly, high levels of missing information might be suggestive of problems with the questionnaire. However, where the amount of missing information varies between interviewers, it could equally suggest that interviewers are not probing sufficiently for answers or are not asking the question properly. Furthermore, there was some concern that recording some information might lead to perverse incentives e.g. if we record the number of times that questionnaires are redistributed to interviewers, researchers might be reluctant to do this even when there is a problem with the questionnaire, for fear it might reflect badly on them.

In order to alleviate their concerns, we are writing accompanying documentation which will give advice as to how these quality indicators should be interpreted and outline their limitations. The quality indicators will initially only be available to project managers and SSD directors, although, in time it is planned that a standard package of quality indicators will be available to clients.

## **8. CONCLUSION**

This project aims to produce a series of quality indicators which will give project managers greater knowledge about what is happening on their surveys but it is also planned that some of these will be released to client organisations. In this way clients will have the necessary information to take account of the quality of all the processes in that are required a survey, rather than just looking at response rates.

## REFERENCES

- Groves M (1991), Measurement Errors Across the Disciplines, in P Biemer et al (eds), *Measurement Errors in Surveys*, Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons, Pages 1-25
- Peter Lynn, National Centre for Social Research, U.K. Johanna Laiho, Statistics Finland, Jean Martin and Roeland Beerten Office for National Statistics, UK, (2001) 'Response rates as a measure of survey quality'. This paper is being presented in session 38 of this conference.