5 Pilot Testing the SHARE Quality Profile

Kirsten Alcser, Grant Benson, and Heidi Guyer

5.1 Background

Quality profiles are gaining increased prominence in survey research. Generally, quality profiles provide a summary of what is known regarding the quality of a survey. They consist of objective assessments of all potential sources for error, including the magnitude of each source. In principle, these data have utility for users of the data and also inform design decisions for subsequent survey implementation, with the acknowledgement that not only is there often a trade-off between cost and quality, but there may also be a trade-off between different error sources.

While there is no commonly accepted standard for what should be included in a quality profile, there is general consensus of key areas among the published reports:

- Statement of core research purposes
- Sample design
- Coverage and sampling error
- Data collection procedures
- Non-response error
- Statistical adjustment procedures
- External comparison
- Documentation and data accessibility

By documenting processes and procedures at each stage of the survey life cycle, potential error is accounted for, starting with sampling and questionnaire design, through data collection, and ending with statistical adjustment procedures and documentation.

Traditionally, quality profiles focus on 'fitness for use' by providing parameters for data generalizability as well as highlighting any potential data pitfalls. As such, the emphasis tends to be on pre-production activities such as sampling and questionnaire design, as well as post-production activities including data cleaning and assigning of sample weights. Much less attention is usually paid to the data collection (i.e. production) process. Thus, it has been virtually impossible for research analysts to assess whether divergence in data collection process and procedure, either across time or data collection agencies, might be at least partially responsible for observed variance in the data.

This chapter reports on the efforts of establishing a production quality profile for SHARE, the Survey of Health, Ageing and Retirement in Europe, which aims at installing this feature as one quality element in the survey. We focus here on the process of setting up the necessary quality monitoring infrastructure, while results from the quality profile will be presented elsewhere.

5.2 The SHARELIFE production quality profile

Starting with the 2008-2009 data collection, SHARE sought to document performance outcome measures collected prior to the post production phases. Collecting and documenting results of quality assessment performance outcomes during preproduction and production, as well as at the end of production, would contribute important contextual information for the benefit of future analysis of the SHARE data as well as serve as an opportunity for survey agencies to make improvements on current and future waves of data collection. This was to be accomplished by providing an on-going feedback and reinforcement mechanism between the coordinating centre (MEA) and the survey data collection agencies. At the conclusion of the pre-test and the main study pre-production phases, survey agencies were provided feedback on adherence to project goals. During data collection, survey agencies were not only provided on-going reports on overall data collection issues, but were also required to submit biweekly status reports on key process indices as a reminder and reinforcement mechanism (see also Chapter 4).

Thus, the first SHARE production quality profile deliberately focused on the assessment and improvement of those aspects of the project that would, in fact, promote *ex-ante* harmonization of the survey process. The critical aspects for the production quality profile were considered to be interviewer recruitment and training, pre-test and main data collection.

The quality profile was first implemented for the SHARELIFE data collection as part of a pilot effort with the University of Michigan Survey Research Center (SRC) consultant team. This was considered a pilot quality profile for several reasons. Significantly, it was unclear at the outset whether the contractual process requirements would have similar effects on data quality for all the countries. Moreover, given that the quality profile data collection instruments were administered in English only to survey agencies with a broad range of experience levels and national traditions, it was uncertain whether technical terms would be interpreted similarly across survey agencies. Thus, the pilot production quality profile data collection was initiated with the expectation that (1) improvements in data collection would be accomplished based on lessons learned in the pilot production quality profile, and (2) a careful review would be undertaken of the quality profile objectives to further establish a common understanding of measurements before formalizing this quality monitoring and documentation tool for future waves of SHARE.

5.3 Implementing the SHARE production quality profile

The SHARE production quality profile was based on the accumulation of production-related measures obtained from and by participating SHARE survey agencies starting with the pretest and ending with main study data collection.

The purpose of the pilot production quality profile was primarily to track survey agency adherence to contractual and quality standards, and to propose data

collection process improvements. In order to accomplish these objectives, information was collected and analyzed from multiple sources, including:

– Agency self-reports
– In-person observations
– Sample management data.

Six primary contractual requirements pertaining to production/data collection were evaluated, which we list in the following. Quotations are taken from the survey agency contracts.

Interviewer Recruitment
The SHARELIFE contract did not provide specific guidelines for interviewer recruitment, but stipulates that each survey agency "shall make sure that as many interviewers as possible who already participated in the SHARE 2004 or 2006 study will also participate in the 2008 SHARE survey." In addition, the contract stated that "All interviewers shall have extensive general face-to-face interview experience. All interviewers personally receive general interview training from [the national] survey agency prior to attending study-specific training. This includes techniques for approaching a household, addressing respondent concerns, probing, recording responses, etc."

Interviewer Training
"Conduct study-specific in-person training sessions of interviewers in local language, each considering the schedule of the "train-the-trainer" programme using materials provided by SHARE" for each stage of the survey process. Additionally, the survey agencies were "responsible for translating interviewer-training materials into local language. Only materials approved by SHARE will be used in the interviewer training sessions." Local training was required to adhere to the content and time specified in the TTT model training agenda. For this wave of data collection, the TTT model agenda devoted approximately 15.5 hours to cover the minimum SHARE-specific General Interviewing Techniques and the SHARE study-specific training.

Interviewers and Number of Interviews per Interviewer
"A minimum of 50 interviewers [are required to] work on the longitudinal main survey. In the main survey, the number of interviews per interviewer must not exceed 50." The primary purpose of these requirements was to limit the effect that interviewers exert on the variance of sample estimates.

Interviewer Effort
This requirement referred specifically to the number of contact attempts: "[A] minimum of eight contact attempts at various days of the week and times of a day is required before a household can be considered as unreachable." Previous SHARE data demonstrate the impact of number of attempts on differential participation rates by demographic characteristics. In order to increase data

consistency between and within survey agencies, the minimum threshold of eight attempts was required prior to coding out a case as a final non-interview.

Retention Rate

While the SHARE contract did not mandate a specific retention rate, it did estimate "fieldwork costs … to achieve a retention rate among subsamples A and B", where subsamples A and B were defined by respondents who participated in wave 2 of data collection (subsample A) or who participated in wave 1 only, but lived in a household where another respondent participated in wave 2 (subsample B). High retention rates were considered critical to the long-term viability of the panel.

Quality Assurance

The SHARE contract required of survey agencies that they "certify that a minimum of 20 percent of each interviewer's completed interviews are verified by supervisory personnel. […] Verification involves calling the respondent by telephone and re-asking factual questions from various parts of the interview. Written records of the verification process must be maintained by <survey agency>." Verification calling is primarily a tool for assuring that interviews were conducted with selected respondents, although such calls may also be used to assess non-response error or data entry error.

5.4 Some first results

Based on these contractual requirements, several on-line surveys were designed to be answered by the agencies before, during and after the fieldwork. Survey agencies completed the survey, capturing, among other things, key outcomes pertaining to specific requirements outlined in the SHARE contract.

- Interviewer training – fulfilment of hours and content covered, as proscribed in the model training agenda. Six survey agencies completely satisfied the requirement across both General Interviewing Techniques (GIT) training requirements and study-specific training requirements, two survey agencies came close, and 3 survey agencies did not meet this requirement.
- Number of interviewers – must recruit minimally 50 interviewers to work on this wave of data collection. Eleven survey agencies met this requirement, while three did not.
- Number of interviews completed by an interviewer – not to exceed a total of 50 interviews. While only two survey agencies met this requirement, overall only a few of the more than 1,000 interviewers did more than 50 interviews.
- Average attempts before coding out as final no contact – a minimum of eight (household) attempts. Only two survey agencies met this requirement.
- Retention rate – a target retention rate of 80 percent for subsamples A and B. Five agencies met the goal, while nine agencies did not.
- Verification call-back – a minimum of 20 percent of all interviews had to be selected for verification. Ten agencies met the verification requirement.

The last results must be seen in the context of the governance of survey work in most European countries: interviewers are usually self-employed with a high degree of independence working for several survey agencies at the same time, diluting the control which SHARE can exert on interviewer work.

In addition to collecting information via the on-line surveys, observations of local trainings took place between the train-the-trainer (TTT) sessions and the start of the fieldwork in the member countries. A total of nine local training sessions were observed between October 19 and October 31, 2008. The consultants from the University of Michigan Survey Research Center observed these local training sessions. Each observation lasted 6-8 hours. Additionally, the SRC observers reviewed the training agenda with the survey agency trainers and the Country Team Leaders (CTLs) to determine whether the content satisfied the SHARELIFE training requirements. Information was collected in a standardized fashion, using a training observation form. This form was divided into three sections: Content review, logistics, and materials.

Interviews were conducted by telephone with selected staff at survey agencies where it was not feasible to conduct observations due to training schedule conflicts (e.g., countries conducting training on the same days). The telephone training debriefing made use of the same training observation assessment form. However, the observations based on telephone interviews were not formally scored and are not included in results reported below.

Training Content & Length of Training
Measurement areas included length of training and content, incorporating a full mock interview (i.e. group practice of the questionnaire administration) and reinforcement of GIT. Six agencies met all training requirements, one agency came close, and two agencies did not meet the requirements for training content.

Logistics
Measurement areas included location, conference setup (e.g. seating arrangements, breakout space as needed); the handling of questions; trainers; presence and participation by the Country Team Leader or Country Team Operator (CTO). All agencies met the requirements, providing an adequate to very good setting for the SHARELIFE training. Also, the request by SHARELIFE for the CTL and/or the CTO to be present and participate in the training as appropriate was met for all but one observed training session.

Training Materials
Measurement areas included training agenda, Interviewer Project Manual, and all other training materials specified in the SHARELIFE TTT programme. Seven agencies met all of these requirements, and two agencies did not meet these requirements. Thus, most survey agencies translated materials provided by the SHARELIFE TTT programme and used these materials to conduct training: PowerPoint slides, SHARELIFE manuals, Sample Management System (SMS) exercises, Grip Strength card, Blaise key card and probe card, etc. However, it was observed that laptops were not available for all interviewers at all trainings.

Most (6 of 9) survey agencies observed by the consultants/trainers met SHARELIFE requirements on all dimensions. Two organizations did not meet the requirements, and one only marginally met requirements.

5.5 Concluding remarks

This was the first step by SHARE to improve the transparency of its complex multi-national fieldwork process. The quality profile project in SHARE collected information about the data collection process and outcomes, producing a production quality profile at the survey agency level as well as for the project overall. Attempts were made to collect information on performance measures based on fulfilment of contractual stipulations, as well as on indicators of quality assurance during the survey process. The emphasis on the outcome of the quality profile documentation for this wave of data collection was to assess agency adherence to contractual requirements, but with a focus on suggested improvements for future data collection efforts.

Because SHARE emphasizes the importance of a standardized training approach for *ex-ante* harmonization of cross-national data collection, it incorporated the direct observation of local training as a component of its evaluation of local training. Consultants observed training in nine of the SHARELIFE survey agencies and conducted telephone interviews with the remaining agencies. The in-person observations proved an invaluable tool in determining root causes for potential deviation from specified project objectives. As outsiders, the observers were able to provide a dispassionate assessment of adherence to project requirements. The goal for future SHARE training evaluation would be to attempt to observe training in all countries. This will require some measure of project level coordination of the schedule of local trainings, so as to avoid overlap of training dates, or an expansion of the consultant observer team to cover multiple local training observations at the same time.

Most agencies successfully recruited interviewers with previous SHARE experience. In fact, it was reported that these interviewers self-selected for work on this wave of data collection. It is highly probable that the proportion of SHARE experienced interviewers will increase with future repeated waves of data collection.

However, in some instances, survey agencies shortened the training because they felt that experienced interviewers already knew how to administer the SHARE instruments and protocols. Thus, it was felt that interviewers who had previous experience with SHARE data collection or who were experienced as a result of working on other survey studies required less training.

There is some concern about retention rate and sample maintenance across waves of data collection (see also Chapter 7). Retention rates of sample A (i.e. respondents who participated in wave II of data collection) were below the required rate of 80 percent for approximately half of the data collection efforts. Lower rates predict a serious decline in the panel sample base moving forward, which is an obvious concern for the project.

For future SHARE Quality Profiles, there are some recommendations regarding the process. First of all, terms used in the online quality profile surveys should be explained in greater detail. More discussion is necessary with members of the survey agencies who coordinate data collection so as to arrive at consensus about dimensions of quality to be measured and captured. Finally since the assessment of quality at the agency level is only as good as the quality of the data entered, survey agencies need to make sure that those knowledgeable of the process provide the information and that it be provided completely.

In conclusion, the main objective for the pilot production quality profile was to identify the types of data that could be collected across agencies to assess the quality of the data collection effort and to identify areas needing improvements. We believe that this has been accomplished.