Cognitive Aspects of Dependent Verification in Survey Operations

Lars Lyberg[1], Stephanie Eckman[2], Magnus Roos[3], Frauke Kreuter[2,4]

[1]Stockholm University, Universitetsvägen 10B, Frescati, Stockholm, Sweden
[2]Institute for Employment Research, Weddigenstr. 20, Nuremberg, Germany
[3]Gothenburg School of Economics, Stockholm, Sweden
[4]Joint Program in Survey Methodology, University of Maryland, 1218 Lefrak Hall, College Park, MD, USA

Abstract
Dependent verification is a commonly used approach to check the validity of survey operations such as coding, listing, and translation. In this context, dependent means that the verifier has access to the initial outcome of the operation, which he/she then reviews for errors. There is ample evidence that this method tends to discover only a fraction of the errors in the initial operation. The authors' experiments with dependent listing and coding verification have shown that listers add only 81% of the housing units that are missing from the initial frame and control coders miss at least 50% of the errors.

Survey methodologists have dealt with this shortcoming in dependent verification by turning to independent verification, where the verifier has no access to the initial outcome. This approach reduces the problem of under detection of errors, but is more expensive and more complex to administer than dependent verification. In addition, independent verification is not itself error free.

By shedding more light on the cognitive mechanisms at work in the verification task, we hope to be able to design more effective dependent verification procedures that neutralize its drawbacks.

Key Words: quality control, CASM, survey error rates, data processing, listing, translation

1. Introduction

Every survey operation is a potential error source that can generate variances and biases that contribute to the total survey error. The main error sources are specification error, sampling error, frame error, nonresponse error, measurement error and data processing error (Biemer and Lyberg 2003). Each source in turn has potential types of error. Examples of error types are concepts used, omission of frame units, item nonresponse, interviewer variance, and erroneous coding. There are various ways of controlling these errors. They can be reduced through preventive measures such as training programs and continuous improvement of underlying processes. For some error types it is also possible to get a reduction through some verification process. Examples of survey processes that lend themselves to verification include coding, editing, listing, keying, and translation. There are two basic types of verification, dependent and independent. In dependent verification a verifier checks a first person's work and decides whether it is correct or not.

If the first person’s work is considered incorrect the verifier changes the outcome. In independent verification two or more persons perform the same work without access to each other’s work. The outcomes are matched and if they differ a decision rule decides the final outcome.

It turns out that the inexpensive dependent verification is inefficient since it corrects only a portion of the errors. The reason is that the verifier is in some way affected by the first person’s work and therefore has a tendency to abstain from making corrections. In independent verification the basis for such problems is removed but the process is much more expensive than dependent verification.

In this paper we will speculate about the nature of the cognitive processes that generate the inefficiency of dependent verification. Ideally we might be able to suggest more efficient dependent verification procedures that neutralize some of the cognitive phenomena associated with current procedures. Thereby, we could benefit from the more favorable cost structure and administrative simplicity relative the more expensive and complicated schemes for independent verification.

In Section 2 and 3 we describe the dependent verification processes used in coding and listing, respectively. In Section 4 we comment on the verification used in editing and translation. Section 5 contains a discussion of what might be going on cognitively in these processes. Section 6, finally, contains suggestions for further work.

2. Coding

Coding is a classification process in which raw survey data, often in the form of responses to open-ended questions, are assigned code numbers or categories that are suitable for estimation, tabulation, and analysis. Coding can be conducted manually by an operator or coder or automatically by specially designed coding software. Sometimes a combination is preferable where the computer codes the easiest situations and the coder codes the remaining ones.

The coding operation has three basic components: A response to a variable for an object, say an individual’s occupation, a coding nomenclature that contains occupation descriptions and associated code numbers, and a coding instruction that relates the response and the nomenclature.

Most coding is susceptible to errors because the coding instructions are not always properly applied and/or the coding instructions themselves are deficient. Developing a quality coding operation is difficult since coding can be a highly subjective activity. Sometimes coders have to use their judgment and read “between the lines” to code the response. Coding operations can be quite large for major surveys and are difficult to manage. Controlling the error in such operations is challenging.

Coding operations can take a number of forms. Coding can be conducted manually at a number of different sites (referred to as manual decentralized coding) or manually in a single site (manual centralized coding). It can also be automated with manual coding of residual cases or the coding can be manual computer-assisted. Variables that typically require coding include industry, occupation, and education. For these variables the nomenclature might contain hundreds of code numbers or categories. We assume that a true code number exists for each element and for each variable under study. A coding error occurs if an element is assigned a code number other than the correct one.

Numerous studies show that the coding error frequencies can be substantial, i.e., the gross errors can be large. For instance, in an RTI study from 1991 described in Biemer and Caspar (1994) estimated error rates for industry and occupation coding were 17% and 21%, respectively. Another is reported in Campanelli, Thomson, Moon, and Staples (1997) describing studies on the quality of occupational coding in the UK. In their studies estimates of correlated coder variance were obtained. The values for ρ were very small and design effects varied between 1 and 1.79 depending on which major occupational group they concerned. A vast majority of the design effects were between 1 and 1.13 showing that the variance inflation factor due to correlated coder variance was relatively modest given workload sizes between 300 and 400. Despite the lack of studies during recent years the problem picture is clear. If coding is left uncontrolled, then error rates are large, which in turn might lead to increased survey error and flawed analyses.

There are basically two different methodologies available for controlling manual coding. These are called dependent verification and independent verification. In dependent verification production coder A codes an element. The code number assigned by A is then reviewed by verifier B. B inspects the code number and decides if it is correct or not. If it is considered correct it remains unchanged; otherwise B changes the code number to one that B thinks is the correct one.

Dependent verification is very inefficient. A common rule of thumb is that only about 20-50 % of the errors are corrected. The cognitive mechanism that creates this low change rate is not really known. It seems as if the verifier’s judgment is strongly influenced by the code number already assigned. The tendency is that only obvious, unequivocal errors are corrected with this method. Less obvious errors tend to remain unchanged. Thus, there is a tendency to defer to the original coder’s judgment.

In independent verification the basis for such doubts on the part of the verifier is removed, i.e., the verifier does not have access to the originally assigned code number. With independent verification, two or more coders independently code an object and after a matching procedure an outgoing code number is decided.

3. Listing

Housing unit listing is a procedure used to create a frame of units for selection for face-to-face surveys. Field staff visit the areas selected for a survey and list all of the residential units that they find there. The lists are returned to the central office where a sample of units is selected. Interviewers then return to the areas to attempt to collect survey data about the residents of the selected households. Sometimes when they list, the field staff start from scratch, with little or no prior information about the housing units in the area. However, when an imperfect address list already exists for an area, the field staff may do dependent listing, in which they update the existing list. They add units that are missing from the list, delete those that are outside of the selected area, are not residential units, or which do not exist. They also confirm the units that are on the list and are appropriate.

Dependent listing suffers from the same confirmation bias discussed above. Eckman (2010) and Eckman & Kreuter (2011) showed that listers fail to add units missing from the initial listing, and fail to delete units on the initial listing that should not be there. These lister errors lead to errors of overcoverage and undercoverage on the housing unit frame, which can result in bias in the collected survey data.

To estimate the size of such effects, Eckman & Kreuter (2011) manipulated the input list given to trained students who acted as listers. This experiment found that units deleted from the units removed from the input list were 15 percentage points less likely to be on the final frame than those that were not removed. Eckman (2010) repeated this experiment in a national study using professional listers, and found effects of a similar magnitude. In addition, this larger study found that the effects were stronger when the removed units were in multi-unit buildings.

4. Editing and translation

Editing is a set of methodologies verifying that survey data captured responses are plausible and, if not, correcting the data. Editing rules can be developed for each variable or for combinations of variables. The editing rules specify likely variable values or likely values of combinations of variables, often as acceptable value intervals. Typically the editing reveals obvious errors and responses highly suspicious of being erroneous. Measures are taken to correct some of them. Sometimes missing values are inserted or recorded values are corrected after recontacting interviewers or respondents. In other cases values are inserted or changed by means of deducing the correct value based upon other information on the record. Most editing is performed automatically by specially designed computer software.

When editing is performed manually there is a risk that it has the same weaknesses as similar dependent verification methods used in other survey processes. The effects of the editing are not well researched. One finding is that there seems to be a tendency towards overediting, i.e., it is very costly and it does not affect the estimates significantly (Kovar and Granquist 1997). Extensive editing is time-consuming and may delay the release of the survey data, thus reducing its relevance to data users. The consensus of the survey community seems to be that the amount of editing should be based upon cost-error optimization strategies. There are many alternative options for reducing survey error and the amount of resources to devote to editing should be balanced against these other options, especially since there is evidence that many editing systems do not improve data quality appreciably. Thus, one way to achieve better results would be to improve the editing procedure itself. Recently editor debriefings have been conducted at Statistics Sweden.

Editor debriefings are a qualitative method where people who work with data editing in a specific survey gather to discuss and report their experiences. A debriefing is similar to a focus group.

The aim is to identify problematic questions and other root causes of errors in the data collection (Svensson 2012). Not surprisingly, it turns out that the editing staff has very good ideas about data collection problems and how they occur. But the anatomy of their work, how they identify a suspicious or erroneous data item, if and how they contact respondents and interviewers, and whether they change responses or not, is not well understood. If it were, this dependent verification could perhaps be adjusted to become more efficient.

Translation of questions and other survey materials is another process usually verified (if at all) in a dependent fashion. The most common method is that translated work is checked by someone. A more sophisticated method is backtranslation, which during recent years has been questioned since it is not completely reliable. Recently, team translation has been suggested. Lots of people are involved in this approach: translators, reviewers, adjudicators, and pretesters. The end product is documented and the process acronym is TRAPD (Harkness 2010). TRAPD is supposed to generate more accurate translations than other methods. But there is a lot of dependent verification associated with both the simplest form of translation verification and TRAPD. There is a lack of more extensive studies of how well these verification systems actually work.

5. Cognitive aspects of dependent verification

The general situation is that dependent verification is inexpensive but ineffective. When this first was discovered in coding processes, the cognitive explanations were limited to phrases such as “the verifier has a tendency to become affected by what has already been assigned” (Minton 1969). When the CASM movement developed in the 1980s it mainly concerned the data collection process. Cognitive aspects of processes such as listing, translation, coding and editing were never investigated.

The obvious remedy was to replace dependent verification with various forms of independent verification. Obviously under an independent scheme a verifier cannot be “affected” by previously assigned values. The problems with independent schemes are twofold. First, they are more expensive. Second, they are not foolproof, due to training and instruction deficiencies. No one has really made an attempt at improving the dependent verification per se. When the ineffectiveness was first discovered actions as the following could have been taken: Those in charge of dependent verification could have gone through actual cases with all staff providing explanations of quality problems and error structures and inviting discussions on how to improve. Experiments on incentives and suitable profiles of the verifiers and the verified could have been launched. Perhaps we would have been able to develop better dependent verification processes and save some resources that are now used for rather elaborate independent schemes.

Now, when we know more about cognition we might be able to speculate what is going on in a dependent verification situation. Four principles of the human mind can serve as guiding principle in this endeavor. The principles need not be mutually exclusive. Our ultimate goal is to conduct some studies or experiments that can suggest improved ways of conducting dependent verification.

Principle 1 says that the human being is striving toward reducing the working capacity of the mind (Cialdini 2001). Let us formulate a hypothesis based on this principle with a bearing on dependent verification. Hypothesis 1: All errors will not be corrected because the verifier is not using his/her mind in a proper and optimal way. This might be related to laziness, lack of incentives and or an overloaded working memory.

Principle 2 says that decisions are made based on heuristics (quick decision rules) rather than reasoned action (Cialdini 2001, Norman 2004, Kahneman 2011). One corresponding hypothesis 2 could be: all errors will not be corrected because the verifier has developed heuristics (gut-feelings, habits, rules of thumb) based on associations to positive and negative emotions from past experiences. On a preconscious level, negative emotions might, for instance, be associated with time wasted searching for nonexistent errors, while positive emotions might be associated with rewards for making quick decisions.

Principle 3 says that the cognitive ability to reflect and reason is subordinated psychosocial mechanisms (such as social conformity and blind obedience) (Asch 1958, Milgram 1974, Cialdini 2001). Hypothesis 3 could be: All errors will not be corrected due to liking, similarity, competence and or authority.

Principle 4 says that the human being is trained to follow logical reasoning rather than to discover unpredictability and randomness (Ariely 2010). Hypothesis 4 could be: All errors will not be corrected because the verifier emphasizes the logic in the system used by the first coder/lister/translator etc rather than paying attention to unpredictable errors.

6. Can current dependent verification be modified to become more efficient?

The next step is to design and launch some studies and experiments that can help suggest improved procedures using dependent verification. At this point it is not quite clear what studies to prioritize.

References

Ariely, D. (2010). *Predictably irrational, the hidden forces that shape our decisions*. New York: Harper.

Asch, S. E. (1958). Effects of group pressure upon modification and distortion of judgments. In Maccoby, E. E., Newcomb, T. M. & Harkness, J. (2010). Survey translation and the question of quality. Paper presented at the workshop on Comparative Survey Design and Implementation, Lausanne, Switzerland, March.

Biemer, P. and Lyberg, L. (2003). *Introduction to survey quality*. Wiley.

Cialdini, R. B. (2001). *Influence, science and practice*. Boston: Allyn & Bacon.

Eckman, S., Kreuter, F. (2011). “Confirmation bias in housing unit listing” *Public Opinion Quarterly*, Vol. 75, No. 1, pp. 139-150.

Eckman, S. (2010). “Errors in Housing Unit Listing and their Effects on Survey Estimates” Ph. D. Dissertation Thesis, University of Maryland.

Granquist, L. and Kovar, J. (1997). “Editing of survey data: How much is enough?” In Lyberg, L. et al (eds). Survey Measurement and Process Quality. Wiley.

Hartley, D. L. (Eds.). *Readings in social psychology* (3rd ed.). New York: Holt, Rinehart & Winston.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books.

Lyberg, L. (1981). “Control of coding operations in statistical investigations. Some Contributions” Ph. D. Dissertation Thesis, Stockholm University, Sweden.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

Norman, D. A. (2004). *Emotional design. Why we love (or hate) everyday things*. Basic Books: New York.

Svensson, J. (2012). “Editing staff debriefings at Statistics Sweden” Paper presented at the Conference for European Statisticians, UN, Oslo, Norway, September 24-26.