

PROCEDURES FOR REDUCING MEASUREMENT ERROR IN ESTABLISHMENT SURVEYS

Karen Goldenberg, Bureau of Labor Statistics; Kerry Levin, Tracey Hagerty, Ted Shen, David Cantor, Westat
Karen L. Goldenberg, Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Room 4985, Washington DC 20212

KEY WORDS: Pretesting, Establishment Survey, Screening, Questionnaire Design¹

1. Measurement Error in Establishment Surveys

In this paper we discuss two sets of procedures designed to reduce measurement error during two separate phases of an establishment survey. The first set is concerned with questionnaire design and pretesting. The second set focuses on steps to identify correct respondents and to contact them for the survey.

Survey measurement error is generally defined as the difference between the answer presented in a survey situation and the true value that applies to that answer. It can originate with the respondent, the mode of data collection, the data collection instrument, and the interviewer (Groves, 1989). In establishment surveys, the establishment's recordkeeping or information system may also contribute (Edwards and Cantor, 1991).

The term *establishment survey* refers to data collection efforts designed to study the characteristics or attributes of organizational entities such as businesses, farms, and institutions (Cox and Chinnappa, 1995). Information sought in such surveys is often record-based administrative and financial data. Therefore researchers usually designate as the respondent a person with specialized knowledge about the survey topic. *Locating* the correct respondent—one who understands the survey questions and has knowledge of and access to relevant records and other information—becomes critically important (Edwards and Cantor, 1991).

Once the designated respondent has been identified and has agreed to participate in the survey, other sources of error, most not unique to establishment surveys, come into play. Common sources of error include whether the correct variables are being measured, the validity of those measures, whether the instrument items mean to the respondent what was intended by the researcher, and completeness and adequacy of the response options. Of course, these sources of error are interdependent. As an example, a poorly-worded question can cause a respondent to provide an incorrect, incomplete, or inadequate answer, and at the same time cause the interviewer to deviate from the structured interview script to assist the respondent. In

establishment surveys, if the information system cannot easily satisfy an information request, a respondent may provide an incorrect answer—or no answer.

2. Background: The Business Births Pilot Study

This paper describes research performed during the Business Births Pilot Study (BBPS), conducted by Westat, Inc. under contract to the U.S. Bureau of Labor Statistics (BLS). The BBPS was created within the context of BLS' Current Employment Statistics (CES) Survey, the source of monthly data on changes in U.S. nonfarm payroll employment. BLS conducts the CES each month with a sample of approximately 400,000 business establishments. While the sample is large, it does not currently include businesses in the early months of their existence. The BBPS was designed to assess whether we could identify these new businesses, called business births, by means of a short telephone interview with the owner or manager of the business, and collect employment data from them.

2.1 Business Births

For purposes of this research, a *business birth* is an establishment, with employees, which formerly had no chance of being selected in the CES survey. Selection into the CES sample is through an employer's Unemployment Insurance (UI) account. The UI program functions separately within each state, and employers open UI accounts in all states where they have operations. Most states require employers to register for an account within either 90 or 180 days of becoming liable for UI taxes in that state, although compliance with this requirement is not universal. While many new UI accounts go to new businesses, other reasons for obtaining them include changes in ownership, mergers or acquisitions, name changes, and incorporations. In general, a new UI account assigned to a continuing economic venture (such as an ownership change) had a prior probability of selection into CES and is not considered a business birth. However, an established business may become a birth by opening a business unit in a state where it did not previously have an account. Similarly, an organization that moves from one state into another becomes a birth in the new state—and a death in the old one.

2.2 Data collection

The BBPS, a telephone survey with monthly data collection, took place between July and November, 1996. Each month, BLS selected a sample of new UI

¹ Views expressed in this paper are those of the authors and do not reflect policy of the Bureau of Labor Statistics or Westat.

accounts and transferred that sample to Westat. Westat interviewers screened the sample units, sent advance letters, and then contacted the designated respondents to conduct a Computer Assisted Telephone Interview (CATI).

2.3 Sample

BLS asked State UI offices for files of new account recipients. Over the course of pretesting and data collection, we used UI accounts from 6 states. We selected a sample of 250 cases for July data collection and 500 in each subsequent month, with approximately equal numbers of cases per state. Files were stratified on the basis of employer size (less than 50 employees and unknown employment, or 50+ employees), with random selection within strata.

3. Development of the BBPS Questionnaire

The BBPS questionnaire was designed for telephone administration to business establishments. The purpose of the questionnaire was to differentiate between business births and continuing economic units among employers that had recently received new UI accounts, and, for the birth units, to measure employment. In designing the questionnaire, we focused on the reason for obtaining a new UI account. When that reason indicated a birth, we looked at when the business began, employment at the time of the birth, and current employment. "Current employment" later evolved to employment under the CES concept, which is the number of paid employees for the pay period including the 12th of the month. We anticipated that the majority of new UI accounts would go to small businesses, and that the respondent would usually be the establishment's owner or manager.

The discussion below presents our efforts to reduce measurement error in this study by (1) conducting cognitive research to assess respondent understanding of key concepts; (2) iterative paper and pencil pretesting to identify and resolve question wording problems; and (3) behavior coding of the first wave of production interviewing to evaluate respondent behavior vis-à-vis the survey items. Questionnaire development and testing took place in 1995 and early 1996.

3.1 Initial design and testing: Cognitive Research

A BLS research team began questionnaire development by attempting to resolve what we meant by "business birth." It was unclear at the outset whether respondents would be familiar with unemployment insurance accounts and whether they could provide the reason their business had received a new one. There were "mechanical" issues to be resolved, such as how to refer to the business if we were speaking to someone

other than the business owner, as well as whether to use open-ended or forced-choice questions on key items.

The first phase of questionnaire testing consisted of a series of cognitive interviews with recent recipients of new UI accounts. Examples of key outcomes from these interviews include: (a) respondents don't necessarily understand what a UI account is, but can still explain why they got a new one; (b) respondents thought in terms of opening a "new" business whether they started one from scratch or purchased it from someone else; and (c) our proposed list of reasons for new UI accounts had missed a very important reason, "hiring employees for the first time." (See Goldenberg, 1996, for a more detailed discussion of the initial research.)

3.2 The CATI Questionnaire—Iterative Pretesting

Pretesting is a critical part of any survey. A new study should have a minimum of two pretests, one that tests initial wording and one that serves as a dress rehearsal for the study as a whole (Converse and Presser, 1986). A second pretest is also important because replacements for questions that didn't work in the first pretest need to be tested; sometimes the new questions are no better or could even be worse than the originals (Sudman et al., 1996).

We conducted three paper and pencil pretests for the BBPS. While we did not plan *a priori* to conduct three separate pretests, we were aware of the likelihood of changes after the initial pretest, and the need to retest the questionnaire following those changes. We decided on a third pretest because we continued to modify the questionnaire and needed to test those changes. We document the BBPS pretest experience here.

3.2.1 Pretest I. The first pretest took place June 3-4, 1996, and resulted in 26 completed interviews out of 40 sample cases. Two experienced Westat interviewers conducted all pretest interviews, which averaged 7 minutes. To evaluate the pretest data, project staff monitored the interviews, and held an informal debriefing with the interviewers. We also compiled a question-by-question summary of every situation where the interviewer was not able to immediately fit the response into a predefined category, or where our subjective assessment showed that the respondent experienced some difficulty with the question. We describe specific pretest results below.

Introduction. The survey introduction attempted to interest the respondent in the study and described the sampled business unit. In addition, it included a reference to the advance letter (see section 4.3), a confiden-

tality statement, and a request for the respondent's permission to tape the interview. If the respondent did not remember seeing the advance letter, the interviewer read it aloud.

Changes: During the introduction, the research team observed signs of respondent impatience, so we shortened it slightly. For respondents who didn't recall receiving the letter, we allowed the interviewer to fax it instead of reading it aloud. We changed the reference to the specific business unit from its complete address to "on [NAME OF] street," and eliminated a second use of that address.

Reason for new UI account. The questionnaire began with an open-ended question on reasons for the new UI account. All 26 respondents answered the unstructured question. Overall, 18 of the 26 respondents (69%) gave what we considered legitimate reasons, all of which appeared on our predefined list. Eight of the 18 were reasons related to having employees and needing a vehicle for paying taxes, while the remaining 10 cited new businesses, ownership changes, incorporations, or other valid reasons. The remaining 8 respondents (31%) referred to legal requirements or processes or implied that they were merely following directions: "It's the law," "my accountant said so," "we're supposed to."

We followed the unstructured question with a list of reasons why a business might obtain a new UI account. The interviewers read each item, and asked respondents to answer "yes" or "no" to each one. The project staff compiled a summary of all responses to the unstructured "reason" question and all individual reasons for which the respondents said "yes." Four respondents said yes to only one reason. Sixteen respondents gave two or more internally consistent reasons (e.g., started new business and hired employees), while the remaining six respondents gave inconsistent reasons (e.g., purchased a business and opened a new business). These inconsistencies had potentially serious consequences, because the respondent could be directed to the wrong section of the questionnaire.

Changes: First, we eliminated the unstructured question, on the grounds that it did not add new information. Second, we acknowledged that respondents had an even broader definition of "new business" than we had previously recognized. We dealt with this situation by asking two questions, one specifically about *starting* a new business or franchise, and the other about the purchase of a business *from someone else*. If the respondent started a new business, and did not purchase it from another owner, we treated it as a newly-created business. If the respondent bought a business, we asked a series of questions about

purchase/ownership change. The interviewer asked additional questions about the reason for a new account only if the new UI account was for reasons other than opening or purchasing a business.

Third, we modified the list of reasons based on responses received in the pretest. We then structured the list so that the more inclusive reasons came first, and instructed the interviewer to read each item until the respondent answered affirmatively. For example, "incorporation" preceded "name change."

Start date and initial employment. Comprehension was not an issue in these questions, but obtaining accurate information was. In some cases the respondent did not know the answers but could obtain them from another source. In other cases respondents estimated the number of employees.

Changes: We trained interviewers to encourage respondents to check records if possible. In some cases interviewers offered to give respondents time to obtain the data and call back, and to accept an estimate only as a last resort. If the respondent had no idea, we told the interviewer to obtain the name and telephone number for another contact and to terminate the interview.

Current employment. We attempted to obtain employment information for the pay period containing the 12th of the current month, or, if that pay period was not yet over, current employment. Nearly a fifth of the pretest respondents (5 of the 26) had difficulty with questions relating the current pay period to the pay period containing the 12th of the month. The pay period of the 12th is an important economic concept for BLS surveys, and its use relies on what Griffiths and Linacre (1995) call the traditional assumption that business surveys are straightforward, with data derived directly from business records by respondents who have a good understanding of economic concepts. They note that this is not always the case, and that the economic concepts traditionally used in business surveys are not easily understood by respondents" (p. 675). The pretest data support this assertion. To assist in locating the pay period, the questionnaire contained the suggestion that respondents refer to a calendar. The interviewers were fairly successful in using it to help "walk" the respondent through the questions. However, the price was considerable deviation from the script and therefore significant interviewer variance.

Changes: We moved the suggestion that the respondent refer to a calendar to the beginning of the series on pay period questions. We addressed most of the pay period of the 12th problems with interviewer training.

Repetitive wording. Respondents whose businesses have employees were asked for the number of employees during the pay period when those employees first reported to work. The question is deliberately long, so as to incorporate cues telling respondents which employees to include in the count:

The number of employees in a company includes full and part time workers, owners of incorporated businesses, temporary workers, managers, executives, corporate officers, office and clerical workers, and all other paid employees who are covered by unemployment insurance. For the *first pay period* that [COMPANY NAME] had employees, what was the number of *paid* employees?

No more than 3 questions later, we asked a nearly identical question about employment for the pay period of the 12th. The excessive wordiness was uncomfortable to interviewers and seemed to annoy respondents. We were concerned that interviewers might not read the questions as worded, so we rewrote the second question in the series.

3.2.2 Pretests II and III. Both the second and third pretests resulted in some changes to the questionnaire, although there were fewer of them and they were less extensive than the changes from Pretest I. Since some of the same issues appeared in both pretests, we discuss them together.

We conducted Pretest II on June 11 with a sample of 30 cases, randomly selected from the same pre-screened file as the first pretest. The sample yielded 20 completed interviews. Like Pretest I, the second pretest identified new issues, questions, and problems. Pretest III took place June 18-20, using an again-revised questionnaire and an initial sample of 102 cases. It yielded 62 completed interviews, with 9 refusals.

Most of the changes we made in the questionnaire following Pretest I appeared to be successful. Modifications to the questions about reasons for a new UI account made the overall interview smoother, and the average length of interview dropped from roughly 7 minutes to about 5 minutes. At the same time, this pretest revealed a few other problems, both with the questionnaire and more broadly with the survey procedures and datafiles.

Reason for UI account. As noted above, if the respondent did not open a new business or purchase one from another owner, the interviewer read a list of possible reasons. Several times the respondent interrupted during the introduction to the question and revealed the reason right away, but the interviewer still had to read all of the items on the list up to and including the appropriate reason. The interruptions signified that the question still required work.

In addition, during the course of the interview three respondents revealed different reasons for their new UI registrations than the ones offered during their first response, and at least one such change affected the outcome variable (nonbirth instead of birth). The presence of these 'informal conversations' indicated that the interviewers were still forced to deviate from the script in order to obtain appropriate responses.

Changes: We continued to shorten the questions, giving the interviewers fewer words to ask. To reduce the likelihood of interruptions in the stem of the 'other reasons' question, we changed it to a single sentence, which seemed to alleviate the problem of respondent

interruptions. However, in Pretest III there were still 6 cases where informal comments during the interview contradicted the original reason. The interviewers were able to go back and proceed through the correct questionnaire path. Since this was a recurring situation, we decided to train interviewers to backtrack if necessary, and made sure that the CATI operation would permit it.

Who is an employee? During the cognitive research we addressed the issue of whether owners of incorporated businesses should be counted as employees, and determined that they should be counted only if they received pay. Owners of unincorporated businesses are not employees and should not be counted. During Pretest II the owner of an unincorporated business included himself in the employee count. Since we already asked whether the business was incorporated, we added a reminder that the total does not include owners to be used when we interview at unincorporated businesses.

Current pay period/pay period of the 12th. Eight respondents from Pretest II (40%) and eight from Pretest III (13%) had difficulty with these questions. After Pretest II, we modified the question about current pay period start date to ask about the *day the current pay period began*, instead of the *first date* of the current pay period. Given the much lower rate of difficulty in Pretest III, the new wording appears to have helped.

Replacing UI accounts. The question about replacing an existing account worked well as long as it was a recent account. In the case of a renewed liability, however, there was some confusion about whether an account still existed. We changed the word 'existing' to active.

3.3 Behavior Coding

Behavior coding has been used since the 1970s to evaluate interviewer behavior, and more recently to evaluate questions for pretests (Oksenberg et al., 1991). Most behavior coding schemes monitor whether the interviewer asks the question exactly, modifies it without changing the meaning, or modifies it significantly. These schemes also note whether the respondent seems to give an adequate answer, asks for clarification, expresses confusion or impatience, interrupts the interviewer with an answer, doesn't know, or refuses to answer. If interviewers have difficulty asking questions exactly as worded, the problem can be either with interviewer performance or with a question that is difficult to ask. If respondents interrupt the interviewer the question is probably too long or poorly worded (as we saw in the first and second pretests). When respondents ask for

clarification or express confusion, again the question is problematic. Measurement error results unless these difficulties are addressed and corrected.

We used behavior coding to evaluate the questionnaire. Two coders analyzed tapes made of 91 of the 150 completed interviews from initial CATI administration period (July). In order to ensure that the two coders were coding interviews in the same way, 39 of the 91 interviews were coded twice—once by each coder.² The analysis that follows is based on the 91 taped interviews and the coding done by the first coder for that interview. The coders selected interviews to code at random, so this should not skew results.

Throughout the pretests, the pay period including the 12th was a persistent problem area in the questionnaire. The behavior coding suggested that the pay period questions required immediate attention. While most respondents were able to supply an adequate answer to these questions, behavior coding pointed to a high level of confusion. Eighty percent of respondents needed clarification from the interviewers about the day the current pay period began, and 75 percent were confused about the relationship between the current pay period and the pay period containing the 12th of the month. We resolved the issue by letting the CATI system take over. We kept the question about the day the current pay period began and we added a question about the day the current pay period ended. In conjunction with the information from a question about how often the firm paid its employees, CATI was programmed to determine whether the current pay period ended before, coincided with, or began after the pay period of the 12th. The CATI system then displayed the appropriate question and asked for the number of employees.

There was relatively little evidence of difficulty with other areas of the questionnaire. However, the questionnaire has many complex routing skips, and several of the questions were only asked a handful of times, or were not asked in the coded interviews.

4. Identifying and Contacting Respondents

A factor which contributes to a decrease in measurement error and an increase in response rates is identifying a respondent who is knowledgeable about the survey content area. In this study the focus was on payroll and employment, so the logical respondent was the individual responsible for a business' payroll.

² Subsequent review of these 39 interviews revealed that there were systematic differences between the two, particularly in the area of interviewer behavior. A third coder recoded some of the same interviews, with a higher percentage of agreement than between the first two coders. See Levin et al., 1996, Appendix A, for a detailed discussion.

Results from other surveys confirm that identifying the most knowledgeable person by name is important to increasing response rates. Paxson et al. (1995) examined the effect of sending a mail questionnaire to a named respondent. Naming the respondent and making repeated contacts produced some of the highest response rates in the 26 surveys they reviewed. However, Moore and Baxter (1993) used contact names and other Total Design Method (TDM) procedures in a mail survey of businesses, and generally found no increase in response as a result. They did find a name helpful, however, among some classifications of small businesses. This is relevant to the present research, as 95% of the establishments are small businesses.

4.1 Preliminary Screening

BLS and Westat have both found prescreening to be a useful technique for identifying the most appropriate respondent in an establishment survey. Prescreening consists of an initial telephone call to the establishment at the outset of the project. During this phone call, the person who answers the call is asked to verify information about the establishment and to provide the name of the person most knowledgeable about the content of the survey. Even if the screening respondent does not supply the correct name, the named respondent is a starting contact within the business.

In the BBPS, Westat began with the business names and telephone numbers provided by BLS. Interviewers called the sampled establishments and identified themselves as calling from Westat on behalf of the Bureau of Labor Statistics. They explained that the purpose of the call was to confirm the business address for a study that was to be conducted in the next few weeks. The interviewers verified the business's mailing address and telephone number, obtained a fax number if one was available, and asked for the name of the person in the company who was responsible for payroll.

The majority of businesses contacted during screening were small, with fewer than 10 employees. Consequently, the individual who answered the telephone was often the owner, head of payroll, or office manager and could also be the survey respondent. Interviewers had to be prepared to answer questions regarding the upcoming survey.

4.2 Screening Results

The primary purpose of screening is to identify a respondent. However, contacting establishments also allowed us to refine and update the sample file.

Identifying correct respondents. Out of a total of 2,248 establishments in our initial sample, we were able to designate a respondent 78% of the time. We

sent advance letters directly to these respondents. Over the five months of data collection, and for the responding sample units, 92 percent of the completed interviews were with named contacts identified during screening. In the remaining cases, the named respondent referred interviewers to another contact (e.g., payroll processing firm, accountant); 45% of these were at different locations from the original respondent.

Identifying closed businesses. Screening allowed us to identify firms that had gone out of business or had otherwise stopped operation (3 percent of the original sample). We excluded these units from the sample file.

Eliminating file errors. Screening also allowed us to eliminate a small number of cases that should never have been included in the sample file. About 2% of the file contained duplicate records, businesses that did not have a location in the state being sampled, or employers who had hired domestic help and were not eligible.

Updating the file. Overall, about 12% of the sample could not be located. Including cases that were out of business or never answered the phone, over 17% of the sample cases were ineligible for further interviewing.

4.3 Advance Letters

The literature strongly supports the use of letters to respondents in advance of the actual survey. The research evidence shows that they ultimately lead to higher response rates, because they enhance the legitimacy of the study and help to prepare the respondent for the survey. Therefore, once we identified the appropriate respondent, we sent a letter addressed to that person by name. The letter was signed by the Commissioner of Labor Statistics. It described the survey, and encouraged the respondent to participate.

During eligibility screening, we asked respondents for a fax number. Across all months of data collection, approximately 70% of respondents provided them. If we had a fax number, we faxed the advance letter; otherwise, we mailed it.

We included a question about the letter in the survey introduction, and most respondents recalled receiving it. Mail or fax made little difference: 75 percent of the fax recipients, and 77 percent of the mail recipients, recalled the letter. However, the time between receipt of the letter and contact with the interviewer by telephone was no longer than one week.

Did the advance letter have an effect on survey response? We did not test this question experimentally, but we think it must have. Response rates for the study ranged from 72 percent to 87 percent each month, with an overall rate of 85 percent.

Rate calculations are based on eligible respondents located during screening.

5. Summary

Overall, we have discussed a number of procedures used to reduce measurement error. First, initial cognitive pretesting led to substantive rethinking of concepts being measured. Second, the questionnaire was subjected to several rounds of pretesting, and the same conceptual issues emerged. Changes made during pretesting were tested iteratively and most problems appeared to have been resolved. Next, some interviews were subjected to behavior coding in order to identify further difficulties with question wording. Since these results demonstrated the persistence of a long-standing problem, we sought an alternate strategy involving a CATI programming change.

In addition, we discuss procedures used to contact sample units and identify the respondent most knowledgeable about the survey content. These procedures include telephone screening to obtain the respondent's name, telephone number, fax number, and current address, and sending advance letters to the respondent. Results indicate that over 90% of the interviews were conducted with the respondent identified during screening. Screening also served to assist in cleaning the data file. Screening also allowed us to update and refine the sample files.

REFERENCES

- Converse, J. and S. Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire*. Beverly Hills: Sage.
- Cox, B.J. B.N. Chinnappa. 1995. "Unique Features of Business Surveys." In B. Cox et al., eds., *Business Survey Methods*. New York: Wiley, 1-17.
- Edwards, W.S. D. Cantor. 1991. "Toward a Response Model in Establishment Surveys." In P. Biemer et al., *Measurement Errors in Surveys*. New York: Wiley, 211-233.
- Goldenberg, K. 1996. "Using Cognitive Testing in the Design of a Business Survey Questionnaire." *Proceedings of the Section on Survey Research Methods*. American Statistical Association.
- Griffiths, G. S. Linacre. 1995. "Quality Assurance for Business Surveys." In B.G. Cox et al., *Business Survey Methods*. New York: Wiley, 673-690.
- Groves, Robert M. 1989. *Survey Errors and Survey Costs*. Wiley.
- Levin, K., T. Shen, S. Heltemes, T. Hagerty, D. Cantor, K. Goldenberg. 1996. *Business Births: Initial Evaluation of CATI Survey*. Westat report submit to the Bureau of Labor Statistics, August.
- Moore, D. R. Baxter. 1993. "Increasing Mail Questionnaire Completion for Business Populations: The effects of personalization and a telephone followup procedure as elements of the Total Design Method," *Proceedings of the International Conference on Establishment Surveys*. American Statistical Association, 496-502.
- Oksenberg, L., C. Cannell, G. Kalton. 1991. "New Strategies for Pretesting Questionnaires." *Journal of Official Statistics* 7: 349-365.
- Paxson, M.C., D.A. Dillman, J. Tarnai. 1995. "Improving Response to Mail Surveys." In B.G. Cox et al., *Business Survey Methods*. New York: Wiley, 303-316.
- Sudman, S., N.M. Bradburn, N. Schwarz. 1996. *Thinking About Answers*. San Francisco: Jossey-Bass.