Response Error: Correlation between Estimation and Episodic Recall Tasks

Nancy A. Mathiowetz
National Center for Health Services Research
Room 18A-55, 5600 Fishers Lane, Rockville, Maryland 20857

Introduction

The term "error" as used in the field of survey methodology has been defined as "the difference between a survey estimate and the value which is estimated" (Hansen, et al., 1951). Several factors throughout the survey process can contribute to the level of error. These factors include the questionnaire topic and question wording, the interviewer's behavior, the respondent's ability to perform the task adequately, and the quality of the post data collection processes including coding, keying, and analyzing the data. The goal of the present work is not to reiterate a discussion of the various components of total survey error, but rather to focus on one source of error, the respondent, and attempt to understand the reasons for his or her inability to report factual information accurately.

Survey respondents are asked to complete several types of cognitive tasks during the course of an interview. These cognitive processes include understanding a question, retrieving relevant information, making judgements, and formulating a response. Not only are the cognitive processes diverse, but the types of questions are numerous. Within one interview a respondent may be asked factual or opinion questions, open or closed ended questions, and questions which require subjective comparisons or estimation.

Cognitive Processes Related to the Survey Interview

The interview process begins with the asking of a question by the interviewer which must be understood by the respondent. Following comprehension of a question, the second task facing a respondent is the retrieval of relevant information necessary to respond. It has long been accepted in survey research that the better the retrieval mechanism, the question, the more accurate the response.

Even in a perfect world with the best sets of cues possible, people are unable to remember events. Forgetting may be the result of one of several factors. The individual may never have transferred the information to long term memory. Sometimes, entire events are not transferred to long term memory; more frequently details about an event may not be transferred. An individual may also lose the ability to distinguish between similar events. Both the recall of specific events and the estimation of the frequency of a class of events may be affected by events occurring between the original encoding of an event and the retrieval process, resulting in a tainting of the originally stored information. Interference theory suggests that forgetting is a function of both the number and temporal pattern of "related" events in long-term memory.

The inhibition and interference theories also imply that single occurrences of an event are more readily recalled than events which fall within a class of related events. The single event which stands out against the backgroud of an individual's life may be the type of event most likely to be recalled, but as discussed below, may cause difficulties with respect to either temporal placement or when used as a basis for estimation tasks.

Once information has been retrieved, the respondent may still need to integrate several pieces of information to answer a question. Tversky and Kahneman (e.g. 1973) have suggested that rules much less precise than rules of probability (called heuristics) are used to integrate information and make judgments. Kahneman and Tversky have identified three of these heurisitcs: availability, representativeness, and anchoring and adjustment.

When people judge the frequency of an event on the availability and speed with which they can recall an occurrence of the event, they are using the availability heuristic. The strength of association is used as a basis for a judgment of frequency. Based on the availability heuristic, the direction of response error would be predicted to be an overestimation of rare, but salient events, and an underestimation of events which are either frequent and closely spaced (since they are not recalled as distinct) or those events that are widely spaced in time and not salient.

The last interview task concerns the respondent's ability to formulate a response. This may involve decision rules related to a choice for closed-ended items and for open-ended items, articulation of the recalled memories.

Research in the area of nonsampling errors has indicated that respondents do not always report accurately. We know, for example, that respondents often err in the dating of events, with most respondents moving the date of the reported event to a more recent date. We also are aware of patterns of omissions in survey reports; the longer the recall period the higher the probability of errors in reports occurring early during that reference period. However, we know little about the consistency of errors among different types of reporting tasks. Do the same respondents make the same errors? Is response error related to the level of the variable one is attempting to measure? How is the pattern of related events correlated to the level of response error?

Research Design

The data presented in this paper are part of a larger study designed to assess the quality of data obtained in the Panel Study of Income Dynamics (Duncan and Mathiowetz, 1985).

Respondents were selected from the personnel records of an established manufacturing company with several thousand employees. Interviews were conducted by telephone using a questionnaire similar to that used in the Panel Study of Income Dynamics, with 78.3% of the 520 respondents participating.

Methods

Information on unemployment was collected in two ways, requiring different estimation and recall techniques on the part of the respondent. The first sequence of questions involved estimation procedures to account for weeks of work and nonwork during the two calendar years prior to July, 1983. The sequence of questions required that the respondent account for the fifty-two weeks of the year by estimating the number of weeks lost from work due to unemployment, vacation, illness, and strikes.

A separate sequence of questions focused on the respondent's ability to accurately recall specific unemployment episodes. Rather than estimate total unemployment for a given year, the task required that the respondent retrieve specific unemployment episodes from memory and report the month that the spell occurred.

Detailed employee records covering the same reference period permitted measurement of the validity of a respondent's report of both total annual unemployment and each month's employment status. Validating the first set of questions concerning calendar year unemployment provides a means of assessing a respondent's ability to make accurate estimates. Validation of the second set of questions concerning months in which unemployment occurred provides information on the accuracy of episodic recall.

Links between the respondent reports and the company records were based on social security numbers. The initial computer match resulted in unmatched records only for the nonrespondents (excess company records). To verify the success of this match, comparisons were made of company and respondent data for several critical data items. For those cases in which the comparisons were outside a prespecified range, the data were hand checked to verify the links.

One concern with validation data is the existence of error in the record data. There was no way to cross validate the company record data; in the absence of a means to assess the error rates in the record data, I have chosen to ignore it for the purposes of this paper and have credited all discrepancies to respondent error. As a result, to some extent, the analysis presented here overestimates the level of response error.

Research Findings

The focus of the first part of the research is on the respondent's ability to estimate the total number of hours of unemployment for the two calendar years preceding the interview conducted in July, 1983. According to the company records, 26% of the sample were unemployed at some point in 1981 and 37% were unemployed during 1982. Table 1 indicates the simple difference and absolute difference between the unemployment hours as reported by the respondent and as recorded in the company records. Table 1 indicates that the estimation process for both recall periods was actually quite good, with no significant difference found in a simple difference test. However, it is interesting to note that a comparison of the absolute difference showed a significant effect for the most recent recall period.

Table 2 provides the same information as Table 1, but only for those respondents with at least some unemployment during either 1981 or 1982. As expected the levels of both interview and record unemployment hours is much higher in this table. However, once again the simple difference in the two reports is not significantly different from zero. The average absolute differences for both 1981 and 1982 are significant, indicating that although over and underreporting do not affect population means, the error that does exist will affect micro-level estimation.

Both tables 1 and 2 indicate that respondents completed the estimation task relatively well. In part, this may be attributed to the context in which the questions were asked. Having the respondent account for all weeks of the year with respect to working, vacation, sick time, and unemployment, may provide an efficient way to have the respondent retrieve this information.

Are respondents consistent in their reporting error across the two years? The correlation between the simple difference in the two years is quite small and insignificant (.03) but the correlation between the absolute error in the two years is significant (.19, $p < .01$). Thus for unemployment, certain respondents are persistently bad reporters but these errors are not consistent in the direction of under or overreporting. It is also interesting to note the correlation between error and the record level of unemployment. The correlation between level and error (simple difference) for 1981 unemployment is -.23 ($p < .01$); for 1982 the correlation is -.13 (N.S.). Given such findings, the assumption used in most measurement error models, that errors have zero means and constant variances, are uncorrelated with each other and with the dependent and independent variables in the models may need to be reexamined.

There is an abundance of evidence in the area of nonsampling errors relating age education, gender, and race to levels of response error. Cannell, et al. (1965) examined each of these demographic factors and found that younger respondents and those with a college education were the most accurate respondents. Other research, both in survey and experimental studies, provides evidence for consistently better reporting among younger, better educated, and white respondents (e.g. Witryol and Kaess, 1957; Loftus, 1979). The effects of gender, however, have been mixed.

Both the simple and absolute differences between interview and record reports of unemployment and the relationship between error and demographic characteristics are examined below. Table 3 presents the analyses for the total sample; Table 4 is limited to those respondents with at least some unemployment in either year. Looking only at Table 3, there appears to be no consistent finding with respect

to demographic characteristics. For example, there are no significant age effects in the simple difference model; the absolute deviation model indicates lower levels of response error for older respondents. This finding is inconsistent with early work which found a persistent relationship between age (older) and higher rates of error.

The analysis on the subgroup with at least some unemployment during either 1981 or 1982 (Table 4) provides evidence that each of the demographic factors is related to levels of response error, although often not in consistent directions for the two years. For example, education is associated with overreporting of unemployment in 1981 and underreporting in 1982. One possible explanation for these changes in the sign of the coefficients may be in the company's unemployment pattern over the two years. In 1982, more people were unemployed than in 1981--respondents may simple be averaging their experience during the two years in their reports.

In contrast to the respondents' relatively accurate reporting of the estimated number of hours of unemployment for each year, their ability to report specific months during which they were unemployed (the specific unemployment spell) was quite poor. Overall, only 35.1% of the spells were reported by the respondents. The proportion reported varied by whether the spell was the first spell experienced by the respondent (34.6% reported accurately) or the most recent spell (49.6% reported accurately), indicating some evidence to support a retroactive inhibition hypotheses for the recall of unemployment experience. The first/last comparison, however, is confounded by the duration of the spells. Duration, as a proxy measure of salience, is believed to be related to reporting, with longer spells hypothesized as having a higher rate of accurate reporting than shorter spells. Among "first" spell, 24% lasted four months or longer as compared to 45% of last spells with a duration of four months or longer.

Table 5 presents logistic regression coefficients for the regression of errors in reports of unemployment. The dependent variable in the model is a dichitomous variable in which a value of "1" indicates that the spell was reported. Looking only at model three, we see that both age and education are significant predictors of accurate reporting. As with the estimation models, older respondents are more likely to report accurately and reports by better educated respondents were more prone to error. Both of these findings are directly contrary to earlier studies of response error.

It is interesting to note in looking at Table 5 that several of the variables used in traditional models of omission errors are significant -- the longer the length of the spell (salience) the more likely it is to be reported accurately and the longer the recall period, the higher the probability of failing to report the spell. However, one of the variables predicted to affect response, interference (as measured by the total number of spells experienced by the person during 1981 and 1982) was not significant in predicting accurate reporting.

The final analysis looks at the relationship between the pattern of events experienced by the respondent, the level of error in estimation tasks and the levels of error in recall of episodic memory. The reader is cautioned that these findings are preliminary and that work is at this time being completed to better understand the relationship between patterns of events and types of response error. Table 6 looks at the levels of error in estimation of unemployment hours for those with only one unemployment spell and those with multiple unemployment spell. As noted above, one would expect that those with only one event would be less accurate in estimating unemployment hours than those with multiple spells. The table indicates the opposite finding--no significant difference in the reports of respondents as compared to company records for those with only one event vs. significant differences for absolute differences among those with multiple spells. The lack of significant differences among those with only one spell may reflect the small cell size (N=44) which limits the power to detect significant differences. The record report of 534 unemployment hours in 1982 among those with only one spell, may indicate that the nature of unemployment spells was quite different for one spell vs. multiple spell respondents, with "one-spell" respondents having one long spell vs. several shorter spells.

Table 7 evaluates the relative importance of demographic characteristics, spell characteristics and accuracy in reporting spell level information in predicting error in estimation. As with the earlier tables, the effects of demographic characteristics is not consistent either across years of across the two measures of error. Both education and gender effect reporting error in 1982 (with higher eduction resulting in higher response error). Characteristics of the pattern of unemployment spells also have mixed effects. The total number of spells, to some extent indicating the difficulty of the reporting task for the respondent, only affects 1982 reports, indicating that the greater the number of spells, the more likely the respondent is to underreport the total number of unemployment hours (simple difference) and the more likely the respondent is to report erroneously (absolute difference). However, total number of spells does not affect error in reports for 1981; rather the mean duration of spells and the mean length of time between the spells and the interview are predictive of 1981 error. This finding may indicate that reference period (for estimates of unemployment hours) may only be important for very long recall periods. Similarly, relative salience (as measured by duration of the spell) appears to be more important in response error models for more distant events.

The final variable presented in each of the models is the proportion of spells the respondent reported accurately. The findings are mixed and difficult to interpret. Correlations between the simple and absolute measures of error and the proportion of spells reported by the respondent indicate a consistent relationship for 1981 and no relationship for 1982 reports.

Conclusions

The data indicate some relationship between errors in estimation and errors in reports of specific unemployment spells. For both sets of tasks, we see that education is related to error, with better educated respondents prone to less accurate responses. With respect to this questionnaire, it appears that the estimation task was far easier for the respondent to complete than the recall of individual events. In part, this difference in ability to report may be related to the design of the questionnaire.

The data also begin to examine the patterns of response error in light of the research completed in the field of cognitive psychology. We see some indication that recall of episodic memory follows a retroactive rather than proactive inhibition models. The data also indicate that hypotheses concerning interference may be useful in understanding error in reporting events or in estimating frequency.

The views expressed in this paper are those of the author, and no official endorsement by the Department of health and Human Services, or the National Center for Health Serevices Research is intended or inferred.

References

Cannell, C. G. Fisher, and T. Bakker, Reporting Hospitalizations in the Health Interview Survey. Washington, D.C.: PHS Vital and Health Statistics, Series 2, No. 6, 1965.

Duncan, G. and N. Mathiowetz, A Validation Study of Economic Survey Data. Ann Arbor, MI: Survey Research Center, 1985.

Hansen, M.H., W. Hurwitz, and M. Bershad, "Measurement Errors in Census and Surveys" Bulletin of the ISI, 38(2):351-374, 1951.

Loftus, E. Eyewitness Testimony. Cambridge, MA: Harvard University Press, 1979.

Tversky, A. and D. Kahneman, "Availability: A Heuristic for Judging Frequency and Probability" Cognitive Psychology, 5:207-232, 1973.

Witryol, E. and D. Kaess, "Sex Differences in Social Memory Tasks", Journal of Abnormal and Social Psychology, 54:343-346, 1957.

Table 1. Response Error in Estimates of Annual Unemployment: Total Sample[1]

	Interview	Record	Simple Difference[2]	Absolute Difference[3]
1982 Unemployment	169 (332)	189 (382)	12	52*
1981 Unemployment	43 (165)	63 (224)	12	49

[1]Numbers presented in table are means. Standard deviations are presented in parentheses. N=387.

[2]Means of simple differnece do not always equal difference in means due to item nonresponse. H_o: $\mu=0$ (two-tailed test).

[3]H_o: $\mu=0$ (one-tailed test).

*$p \leq .05$

Source: PSID Validation Study

Table 2. Response Error in Estimates of Annual Unemployment: Respondents with Some Unemployment in 1981 or 1982[1]

	Interview	Record	Simple Difference[2]	Absolute Difference[3]
1982 Unemployment	473 (421)	539 (481)	43	140**
1981 Unemployment	106 (216)	181 (353)	52	126**

[1]Numbers presented in table are means. Standard deviations are presented in parentheses. N=133.

[2]Means of simple differnece do not always equal difference in means due to item nonresponse. H_o: $\mu=0$ (two-tailed test).

[3]H_o: $\mu=0$ (one-tailed test).

**$p \leq .01$

Source: PSID Validation Study

Table 3. Coefficients for the Regression of Errors in Reports of Unemployment on Demographic Factors: Total Sample

	Simple Difference[1]		Absolute Difference[2]	
	1981 Unemployment	1982 Unemployment	1981 Unemployment	1982 Unemployment
Age (Years)	0.87	-0.29	-1.78*	-1.97*
Black	-55.84*	18.23	35.02	4.87
Female	-64.18*	-46.81	49.86	31.96
Education	0.50	-9.63*	3.30	7.22*

[1]Simple difference between the interview and record report.

[2]Absolute difference between the interview and record report.

*p ≤ .05

Source: PSID Validation Study

Table 4. Coefficients for the Regression of Errors in Reports of Unemployment on Demographic Factors: Those with Some Unemployment during 1981 or 1982 (N=133)

	Simple Difference[1]		Absolute Difference[2]	
	1981 Unemployment	1982 Unemployment	1981 Unemployment	1982 Unemployment
Age (Years)	3.79*	-1.90	-3.67*	-0.96
Black	-156.57**	14.10	103.08*	58.33*
Female	-100.74**	-64.90*	54.82*	3.34
Education	27.36**	-19.20**	-13.12**	19.35**

[1]Simple difference between the interview and record report.

[2]Absolute difference between the interview and record report.

*p ≤ .05
**p ≤ .01

Source: PSID Validation Study

Table 5. Logistic Regression Coefficients for the Regression of Errors in Reports of Unemployment Spells[1]

	Model I	Model II	Model III
Duration			
1 week	---	---	---
2 weeks	.24	.34	.49
3 weeks	.40	.23	.28
4 weeks	1.17**	1.02	1.20*
5-16 weeks	.77	.71	.91*
17+ weeks	1.36**	1.28**	1.55**
Spell Number		.05	.06
Total Spells		-.00	.00
Recall Length (weeks)		-.01*	-.01*
Age			.03*
Education			-.20**
$\Delta\chi^2$		13.7	17.2
p		.01	.001

[1]Dependent variable is a dichotomous variable indicating whether the unemployment spell was reported by the respondent (1=reported)

*p ≤ .05 **p ≤ .01

Source: PSID Validation Study

Table 6. Response Error in Estimates of Annual Unemployment by Number of Discrete Unemployment Spells[1]

	Interview	Record	Simple Difference[2]	Absolute Difference[3]
Only 1 spell (N=44)				
1982 Unemployment	492 (478)	534 (561)	-25	102
1981 Unemployment	25 (111)	188 (460)	95	113
More than 1 spell (N=87)				
1982 Unemployment	464 (392)	541 (437)	78	159**
1981 Unemployment	147 (244)	178 (285)	30	132**

[1]Numbers presented in table are means. Standard deviations are presented in parentheses. N=387.

[2]Means of simple differnece do not always equal difference in means due to item nonresponse. H_o: $\mu=0$ (two-tailed test).

[3]H_o: $\mu=0$ (one-tailed test).

**$p \leq .01$

Source: PSID Validation Study

Table 7. Coefficients for the Regression of Errors in Reports of Unemployment on Demographic Factors and Spell Characteristics

	Simple Difference		Absolute Difference	
	1981 Unemployment	1982 Unemployment	1981 Unemployment	1982 Unemployment
Age	2.69	-1.81	.72	-4.36*
Education	5.25	24.66*	13.61	33.78**
Sex	180.65*	267.12**	134.94	135.52*
Total Number of Spells	-2.34	22.95**	4.88	13.99**
Mean Duration of Spells	6.29**	.31	4.25**	-.53
Mean Length of Time between Interview and Event	4.42**	-1.28	4.44**	-.64
Proportion of Spells Reported Accurately	-146.21**	-.37	-80.76	71.61*

*$p \leq .05$
**$p \leq .01$

Source: PSID Validation Study