Recall Strategies and Memory for Health-care Visits

JARED B. JOBE,
ANDREW A. WHITE,
CATHERINE L. KELLEY,
DAVID J. MINGAY, and
MARCUS J. SANCHEZ

National Center for Health Statistics

ELIZABETH F. LOFTUS

University of Washington

EPIDEMIOLOGIC AND STATISTICAL DATA FROM national health surveys, such as the National Health Interview Survey (NHIS) (e.g., Kovar and Poe 1985), are used to plan, legislate, administer, and evaluate federal health programs. The accuracy of these data is therefore of the utmost importance both to public policy makers and to scientists. Moreover, the type of information needed both for scientific research and for the formulation and assessment of health and social policies has become increasingly sophisticated. To meet these data requirements, survey questions have become more complex, placing great cognitive demands on survey respondents. Survey questions are often difficult to comprehend, place heavy demands on memory (e.g., "How often have you eaten chicken in the last 12 months?"), and require difficult judgments.

In the last decade particularly, individual researchers (e.g., Loftus

The Milbank Quarterly, Vol. 68, No. 2, 1990

1982) and expert panels (Jabine et al. 1984; Hippler, Schwarz, and Sudman 1987) have argued that the methods and theories of cognitive science, if applied to survey research, could improve the accuracy of survey data. Scientific advances for both cognitive science and survey research were predicted. Spurred by the reports from these expert panels, collaborative research between the two disciplines has begun in earnest (see e.g., Blair and Burton 1987; Blair and Ganesh in press; Jobe and Mingay 1989; Means and Loftus in press; Royston et al. 1986; Smith, Jobe, and Mingay in press; Schwarz, Hippler, and Noelle-Neumann in press).

Jobe and Mingay (in press) describe the benefits that are accruing to each discipline from collaboration. One of these benefits is that cognitive methods and theories can be applied to survey experiments conducted in the field. In this project, we replicated a laboratory experiment from cognitive science using an experiment embedded in a field survey. The subject we selected is of both theoretical and practical relevance: the relationship between the order in which respondents recall events and the accuracy of recall. The practical relevance arises from the need to develop questioning strategies that increase the accuracy of recall from autobiographical (everyday) memory. If research shows that the order in which individuals choose to recall does not result in maximal recall, then respondents might be instructed to use a more effective order.

The theoretical relevance arises from competing predictions of whether a forward (chronological), backward (reverse chronological), or free order (no instruction) should lead to superior reporting. These theoretical perspectives are discussed by Whitten and Leonard (1981) and Loftus and Fathi (1985). Briefly, some theories suggest that forward recall would be better because recall of the earlier events could guide the recall of subsequent events and the recall order matches the sequence in which the events were experienced. However, other theories suggest that backward order may be superior because recent events are likely to be more easily remembered and thus offer better cues for searching memory for related events. Finally, order of retrieval may have no effect on memory performance if these advantages of forward and backward recall balance each other or if the events to be recalled are either completely independent of or fully dependent on one another.

Several earlier studies examining order of recall have yielded inconsistent results (Loftus 1984; Loftus and Fathi 1985; Loftus et al. 1989;

Whitten and Leonard 1981), with no clear pattern emerging as to which method of recall is best. Moreover, it appears that the nature of the to-be-recalled material interacts with recall order. Loftus et al. (1989), in a study of particular relevance to the present investigation, found a nonsignificant trend for backward order to be better than forward or free recall for the recall of medical-provider visits. Respondents were health-maintenance-organization (HMO) subscribers recalling medical visits.

Another issue of theoretical and practical relevance to the present study is the number of events to be recalled. It would be reasonable to assume that respondents with numerous visits would be unable to recall each visit individually (episodic recall), in which case recall order may be an ineffective strategy. This hypothesis is consistent with the results of Means and Loftus (in press, study 3), namely, that HMO subscribers appeared to be unable to recall episodically their medical-provider visits for the same condition when the number of visits in the reference period exceeded four. Other data also support this hypothesis and indicate that respondents use strategies other than episodic recall when answering questions about high-frequency behavior (e.g., Blair and Burton 1987).

Given the inconsistent results from studies investigating order of recall, we designed our experiment to study the effect of recall order on the accuracy of recalling medical-provider visits in a field survey. The issue is important because many health surveys ask respondents to recall a series of health-related events, such as visits to a medical provider. It was of secondary interest whether recall order would be an effective strategy for respondents who had a large number of visits to recall. Understanding the circumstances under which particular questioning procedures facilitate recall should promote understanding of how autobiographical information is organized in long-term memory, thus aiding the development of models of long-term memory.

Method

Sampling Plan

This experiment was conducted as part of the NHIS/National Medical Expenditure Survey (NHIS/NMES) Linkage Field Test. Although the original sample design of the NHIS is a complex probability sample of the civilian, noninstitutionalized United States residents, the sample for the linkage field test was a purposive one and was *not* designed to be representative. Demographic groups tending to have more health problems and less access to health care were oversampled in order to test the feasibility of linkage. Blacks, women, Hispanics, the aged, and the poor composed a larger proportion of the sample than of the overall United States.

The sample frame consisted of the NHIS weekly samples of dwelling units in eight sites for a three-and-a-half-month period. The eight sites were U.S. Bureau of the Census primary sampling units (PSUs) representing counties in and around Baltimore, Chicago, Detroit, Milwaukee, Los Angeles, San Antonio, San Francisco, and Tippecanoe County, Indiana. The sampling plan called for 600 respondents; however, when a respondent moved to another household within the same PSU, that respondent was interviewed in his or her new household, and the new occupants of the old household were interviewed as well. As a result, the final sample consisted of 628 respondents, of whom 19 were declared ineligible (e.g., vacant or student), 34 were not locatable, and 68 refused, leaving 507 completed interviews for the experimental portion of the field test. For a detailed description of the field test, see Mathiowetz and Ward (1987).

Interview Design and Procedure

The NHIS/NMES interviews were conducted face to face in respondents' households. Households were randomly assigned to one of three "order" conditions: forward (chronological) recall, backward (reverse chronological) recall, or free (uninstructed) recall. The "order" question was the first question in the survey interview. Each contractor-trained interviewer administered all conditions and was not informed of the experiment's purpose. The interviewer asked each adult respondent present in the household to recall the month and day of the month for all of his or her medical-provider visits over the previous six months and to give a basic reason for each visit. The interviewer attempted to interview each respondent in private, out of earshot of everyone else in the household, to maximize honesty of reporting and to avoid generating recall cues for subsequent respondents. This precaution, combined with the random assignment of recall-order conditions within households, meant that, even in the unlikely event that recall cues were inadvertently generated for some subsequent respondent, they could not have a differential effect across recall conditions. The questionnaire is shown in figure 1.

Following the interview, the interviewer asked each respondent who reported at least one visit to a medical provider to sign permission forms authorizing us to obtain information regarding his or her medical visits from each named provider.

Medical-provider Contact

Of the 507 respondents with complete interviews, 337 reported having at least one medical-provider visit during the six-month reference period (66 percent). (The recall experiment included only those respondents who reported medical-provider visits.) Contractors attempted to contact all the 542 medical providers named by respondents, requesting information on respondents' reported visits during the reference period. Contractors reviewed all medical records obtained from the medical providers. They compared respondents' reports to the providers' records and attempted to reconcile any visits reported by the respondent that were not found by the provider. The information collected from the recall section of the interview, along with the information collected from the medical providers named by the respondent, were used to verify the reported visit date and reason for visit. Using a very strict definition of a complete medical record, we obtained complete medical records for 130 of the 337 respondents reporting at least one visit, for a 39 percent completion rate. A complete medical record was defined as one in which all medical providers named by the respondent submitted information on visits and visit dates.

The rather large loss in data due to physician nonresponse is attributable to several causes. Physician refusal was low; only 30 out of 542 (6 percent) refused to participate in the project. Twenty-four physicians were retired or deceased, or the practice was closed or not locatable. Thirty-one of the physicians named by respondents could not find records for the relevant respondents. Fifty-five of the providers contacted were not M.D.s or D.O.s, and thus not eligible for inclusion in this study. We considered eliminating these providers and conducting the matching analysis on each respondent's remaining visits. However, it was often difficult to determine which of the respondent-reported visits should be removed. We therefore decided to eliminate these

	Column 1 Person # ____ Name	Column 2 Person # ____ Name
A1. ADMINISTER RECALL SECTION TO EACH ADULT (AGE 18 OR OLDER) PRESENT. USE THE QUESTION FORMAT (A, B, OR C) CHECKED. (RECORD ALL VISITS MENTIONED, IN THE ORDER MENTIONED BY THE RESPONDENT.) To help you start thinking about the things we'll be talking about during the interview, I'd like you to tell me about each time you saw a medical person during the period from (REFERENCE DATE) to (TODAY'S DATE). For each time you saw a medical person, please tell me the date of the visit and the reason you went. (I'd like to start with (you/NAME) and have each of the adults answer separately for themselves.) A. ☐ Please begin with the first visit after (REFERENCE DATE), then tell me about the next visit after that, and so on up to today. (PROBE: And when was the next visit after that?) B. ☐ Please begin with the most recent visit, then tell me about the one before that, and so on, back to (REFERENCE DATE). (PROBE: And when was the next visit before that?) C. ☐ Please begin with any visit you want and tell me about your visits in any order you wish. (PROBE: Were there any any other visits? When were they?)	a. ____ / ____ MM DD REASON:	a. ____ / ____ MM DD REASON:
	b. ____ / ____ MM DD REASON:	b. ____ / ____ MM DD REASON:
	c. ____ / ____ MM DD REASON:	c. ____ / ____ MM DD REASON:
	d. ____ / ____ MM DD REASON:	d. ____ / ____ MM DD REASON:
	(Columns continued to accommodate up to 6 adult self-respondents and rows continued to accommodate up to 15 visits.)	

FIG 1. Sample questionnaire.

respondents and their records from the data set. Records from 67 physicians were not obtained for a variety of other reasons, such as time and procedural limitations imposed by the contract. Of the 542 physicians contacted, 335 medical-provider records were obtained, for a 62 percent return rate.

The discrepancy between the physician return rate and the percent of respondent data lost due to incomplete medical records derives from some respondents visiting several providers. If it is assumed that completion of records by each provider is an independent event, then the completion rate for respondents with one reported provider should be 62 percent; for those with two reported providers this rate should be 38 percent (or $.62^2$); for those with three reported providers the completion rate should be 24 percent ($.61^3$); and so forth. This implies that more data were lost from respondents with visits to several medical providers. Most of our data is for respondents with two through four visits.

Matching Procedure

After determining which respondents had complete health-care-provider information, we matched the recall information with provider information for those respondents ($N = 130$). A recall/record date pair was considered a "match" or "correctly recalled visit" only if the reason given for the visit by the respondent corresponded with the reason given by the medical provider. Determining a reason match was sometimes a complex process. For example, respondents often reported symptoms, whereas providers often reported diagnoses. Nevertheless, relying on our medical knowledge and on medical manuals and dictionaries, we were able to reach accord on reason matches. Furthermore, three strictness criteria were defined on the basis of chronological distance between the recall and record dates: (1) strict (respondent-reported date was within 15 days of provider date); (2) moderate (within 30 days of provider date); and (3) loose (within 60 days of provider date). Because the results of the three strictness criteria were similar, only those for the moderate criteria are reported. Three authors judged initial matching on a subset of the records and at least two authors agreed on classification on 99 percent of the records. Two authors then judged matching on a subset of the records. When 85 percent interrater agreement was obtained for these, one author coded the remaining records.

Two sets of dependent measures were defined for statistical analyses. The variable called "completeness" represents the degree to which a respondent was able to recall accurately all of the visits listed by the providers. Completeness could be considered a measure of underreporting, corrected by removing uncodable or unmatchable visits reported by respondents. It was calculated as correctly recalled visits (CRV)/total visits in the medical records. The second variable, called "accuracy," represents the degree to which the visits recalled by respondents are both codable and matchable. It was calculated as CRV/total visits reported by respondents.

Analysis Plan

Our data analyses addressed several issues. First, we compared the distribution of several important variables in our sample with the 1987 NHIS self-respondents. Second, we examined the extent of underreporting of health-care visits. This analysis was conducted on the basic sample of 130 respondents who reported at least one health-care visit and had complete medical-provider records. Third, we determined the extent to which the respondents followed the recall-order instructions. Fourth, using matching analyses, we examined how recall instruction affected accuracy and completeness of recall. The third and fourth analyses used a subset of 75 respondents for analytic and theoretical reasons discussed below. Next we conducted analyses varying the number of respondent visits, to test the hypothesis that recall instruction would have a greater effect on respondents reporting two to four visits than it would for respondents reporting five or more visits.

After completing these basic, hypothesis-testing analyses, we conducted additional analyses to determine if the secondary variables of respondent age, race, gender, income level, level of education, and self-reported health status had any effect on the recall patterns obtained. These secondary analyses were conducted both to generate new hypotheses and to provide insights into how generalizations about our results might extend to different groups of people. They were not designed to test any specific hypotheses.

Results

Representativeness of the Sample

The sampling plan called for oversampling certain groups, did not specify a probability design, and was not intended to be a representative sample. However, because of the potential for secondary variables to interact with independent variables, we compared the distribution of respondents with important characteristics in our final sample with those in the 1987 NHIS (data for whom were readily available). Whereas women comprised 67 percent of our final sample, they comprised 64 percent of the NHIS self-respondents. Our final sample contained a significantly higher percentage of the oversampled groups (minorities, elderly, and people who considered themselves to be in poorer health) than the NHIS. Table 1 shows the comparisons between our sample and the 1987 NHIS on several important secondary variables.

Underreporting

Results indicated that, overall, respondents underreported the number of medical-provider visits by 20 percent, with a total of 341 visits reported by the 130 respondents who reported one or more doctor visits, and 428 visits reported by the medical providers for these 130 respon-

TABLE 1
Comparison of Final Sample with 1987 NHIS Sample

Characteristic	1987 NHIS (n = 55,472)	Our final sample (n = 75)	χ^2	p
Female	64.0%	67.0%	0.23	n.s.
Minority	16.7%	56.5%[a]	78.56	<.001
Over 65	20.5%	53.0%	47.62	<.001
Reporting excellent/ good health status	86.2%	59.5%[b]	36.67	<.001

[a] n = 69.
[b] n = 74.

dents. This figure is likely to be an underestimate of underreporting for several reasons. First, respondents who were interviewed but did not report any provider visits, even though he or she had made some visits, were not included in any analyses, because if the respondent reported no visits, no medical providers were contacted. Second, if a respondent visited several medical providers in the reference period, it is possible that he or she failed to recall one or more of these providers. Unless a provider's name was reported, it was not possible to contact that provider. This underreporting does not seriously affect the analyses presented below, because differences between experimental groups rather than amount of error are the focus of concern. However, it does affect estimates of underreporting for survey descriptive purposes. Respondent- and provider-reported visits are presented in table 2.

Order of Recall

Respondents with one visit were not included in these analyses because the order-of-recall instructions were not expected to influence recall of a single visit. Thus, 75 respondents were included in these analyses.

Following Instructions. Respondents followed order instructions a majority of the time. Across all respondents in the forward-recall condition, 78 percent of the second or subsequent visits were given a more recent date than the prior reported visit. In the backward-recall condition, 75 percent of the visits were dated as being more distant than the

TABLE 2
Average Number of Visits Reported

	Order of retrieval			
Variable[a]	Forward (n = 43)	Backward (n = 43)	Free recall (n = 44)	Total (n = 130)
Number of provider-reported visits	4.27	4.09	5.22	4.51
Number of respondent-reported visits	3.30	3.91	4.35	3.80

[a] Includes all respondents with one or more reported visits.

prior reported visit. In the free-order condition, there was a preference to move in a forward temporal order, with 77 percent of visits being given a more recent date than the prior reported visit.

Completeness and Accuracy. Analyses of variance were computed, assessing matching errors for completeness and accuracy, for all respondents reporting two or more visits. Results of both analyses are shown in table 3.

For the completeness-of-recall variable, there was no significant effect of recall condition; the tendency for the free-recall condition to result in better performance fell far short of significance (37 percent, 43 percent, and 53 percent for forward, backward, and free recall, $F[2,72] = 1.37$, $p = .26$). For the accuracy-of-recall variable, free recall was superior to either forward or backward recall, with respondents correctly recalling 47 percent and 42 percent of their visits in the forward- and backward-recall conditions, compared with 67 percent of their visits in the free-recall condition. This finding was supported by a marginally significant main effect for recall condition ($F[2,72] = 2.78$, $p = .07$).

Analyses Varying Number of Respondent Visits. To test the hypothesis that any effects due to recall group would be magnified for respondents with a smaller number of visits, the analyses of variance were repeated restricting the sample to those reporting two to four visits (n = 54). This prediction was not supported. Means for accuracy were 43 percent, 44 percent, and 70 percent, for the forward-, backward-,

TABLE 3
Performance of Respondents Reporting Two or More Visits as a Function of Order of Retrieval (in percent)

	Order of retrieval			
Variable	Forward (n = 30)	Backward (n = 22)	Free recall (n = 23)	Total (n = 75)
Completeness of recall[a]	37	43	53	44
Accuracy of recall[b]	47	42	67	52

[a] Completeness of recall = matchable respondent visits/total provider visits $F(2,72) = 1.37$, $p = .26$.
[b] Accuracy of recall = matchable respondent visits/total respondent visits $F(2,72) = 2.78$, $p = .07$.

and free-recall conditions. The main effect for recall condition was again only marginally significant ($F[2,51] = 2.45$, $p < .10$). However, when the analysis was repeated (the two analyses are *not* independent), using only respondents who reported two to five visits, the effects reached statistical significance, with 38 percent correct for both forward and backward recall, compared with 69 percent for free recall ($F[2,57] = 3.66$, $p = .03$).

Secondary Variables

Analyses were then conducted to determine if recall condition interacted with any secondary variables of interest.

Respondent Gender. Using the 75 respondents with two or more visits, for females free recall was superior in completeness to forward or backward recall. Sixty percent of doctor visits were correctly recalled in the free-recall condition, compared with 40 percent and 29 percent of visits in the forward and backward conditions, respectively. For males, 61 percent of doctor visits were correctly recalled in the backward-recall condition, with 31 percent and 38 percent of visits correctly recalled in the forward- and free-recall conditions, respectively. This finding was supported by a significant interaction of recall condition with respondent gender ($F[2,69] = 3.37$, $p = .04$). These data are presented in table 4.

Further analyses were conducted to investigate possible reasons for the interaction. First, the interaction of gender and recall condition could be due to females having more medical visits than males. However, there were no statistically significant differences in the number of provider-reported visits due to gender ($F[1,62] < 1$).

Another possible reason for the gender and recall-group interaction is that females may have multiple reasons for medical visits, whereas males may make multiple visits for one health condition. Free recall might be superior for those with multiple reasons; one visit would act as a retrieval cue for other visits made for the same reason and only the free-recall condition would permit respondents to report in this way. To test this hypothesis, the sample of 75 respondents with two or more reported visits was recoded into two groups: those with one reason for medical visits and those with two or more reasons for medical visits. This grouping was based on physician rather than respondent records because respondents' reports of reasons for visits were often vague and

TABLE 4
Effects of Gender and Order of Retrieval on Completeness of Recall[a]

	Order of retrieval			
Variable	Forward (n = 30)	Backward (n = 22)	Free recall (n = 23)	Total (n = 75)
Males				
Percent	31	61	38	45
Number	8	10	7	25
Females				
Percent	40	29	60	44
Number	22	12	16	50

[a] Completeness of recall = matchable respondent visits/total provider visits
Recall order main effect: $F(2,69) = 1.44$, $p = .24$
Gender main effect: $F(1,69) = 0.02$, $p = .88$
Gender × recall-order interaction: $F(2,69) = 3.37$, $p = .04$.

therefore not codable (uncodable visits would only be included in the previously reported analyses in the denominator of the "accuracy" variable). Because any respondents with fewer than two provider-reported visits were eliminated from this analysis, the sample size dropped slightly, from 75 to 70. Males and females had proportionately similar patterns of medical visits, with 36 percent of males having only one reason for visits versus 64 percent with multiple reasons, and 31 percent of females having only one reason versus 69 percent with multiple reasons. Moreover, number of reasons for medical visits did not interact with recall group on completeness of recall ($F[2,69] = 0.00$, $p = 1$), nor were there any other significant effects involving number of reasons for medical visits. Thus, differences in the number of reasons for visits (i.e., homogeneity) did not explain the gender interaction. The gender effect may only be an artifact.

Self-reported Health Status. A self-reported health-status question used in the interview asked respondents to indicate whether they perceived their health to be excellent, good, fair, or poor. Respondents reporting "excellent" (n = 10) or "good" (n = 34) were considered to regard their health as good for this analysis, whereas those reporting "fair" (n = 20) or "poor" (n = 10) were considered to regard their health as poor for this analysis (one respondent did not answer). An analysis was conducted to determine whether respondents reporting poor health recalled more medical-provider visits than respondents reporting good health. There were no differences in the number of provider-reported visits recalled ($F[1,62] < 1$). However, self-reported health status showed a marginally significant interaction with recall group on the accuracy variable ($F[2,68] = 2.61$, $p = .08$). For those reporting good health, mean accuracy scores were 48 percent, 39 percent, and 85 percent for the forward-, backward-, and free-recall conditions, respectively. For those reporting poor health, mean accuracy scores were 40 percent, 50 percent, and 50 percent for the forward-, backward-, and free-recall conditions, respectively. Health-status data are presented in table 5.

Respondent Age. Analyses were conducted using respondent age (coded into groups of less than 25, 25 through 39, 40 through 59, and 60 or over) as an independent variable. Although a significant interaction of age and recall group was obtained for both accuracy and completeness, these effects are uninterpretable for two reasons: First, the interaction did not become more accentuated as age increased. Second, the cell sizes were widely divergent and in many cases extremely small (six out of twelve cells had four or fewer observations, and one cell had only one observation).

TABLE 5
Effects of Self-Reported Health Status and Order of Retrieval on Accuracy of Recall[a]

Variable	Order of retrieval: Forward (n = 30)	Backward (n = 22)	Free recall (n = 23)	Total (n = 74)[b]
Good health				
Percent	48	39	85	54
Number	17	16	11	44
Poor health				
Percent	40	50	50	46
Number	12	6	12	30

[a] Accuracy of recall = matchable respondent visits/total respondent visits
Recall order main effect: $F(2,68) = 3.14$, $p = .05$
Health-status main effect: $F(1,68) = 0.81$, $p = .37$
Health status × recall order: $F(2,68) = 2.61$, $p = .08$.
[b] One respondent was lost due to nonresponse on the health-status question.

Other Variables. Analyses were also conducted using respondent race, education level, and income as independent variables. None of these variables resulted in significant main effects or interactions.

Discussion

Recall performance of respondents who were allowed to recall medical-provider visits in whatever order they chose was superior to that of respondents instructed to recall in a temporal order. This finding suggests that there are occasions when temporal order is not the most effective recall procedure, especially when all of an individual's visits are for unrelated health conditions. Under these circumstances, the flexibility provided by free-recall instructions offers respondents the chance to recall in the most effective manner, given the number and type of health-provider visits they have made.

Conversely, some of the events that were recalled in the required order in the forward- or backward-order conditions (78 percent and 75 percent, respectively) may have been recalled in a somewhat different order, but were then mentally reordered into the required temporal structure prior to being reported. This extra task may be partly responsible for the poorer performance of respondents in the forward- and backward-order condition compared with the free-recall condition.

The marginal superiority of free recall to forward and backward temporal recall reached significance when respondents reported two to five visits, although this conclusion must be tempered by the fact that the analysis was post hoc. This finding is somewhat consistent with the results of Means and Loftus (in press), who found that memory for medical visits became generic (i.e., they could only remember a typical visit) when five or more visits occurred for a chronic condition, with respondents being unable to recall the individual visits. Differences in the reference period, experimental procedures, or type of health-care services used might account for the small differences in the results of the two studies. In the experiment reported here, too few respondents had five or more visits for detailed analysis. However, it may be that our recall-order instructions are only effective when medical visits are sufficiently infrequent that they are remembered as specific episodes. When the visits are more frequent, memory for them becomes generic.

The interaction between self-reported health status and recall instructions was mainly due to respondents who reported good health also providing accurate dates and reasons foı their visits 85 percent of the time in the free-recall condition, an unusually high recall rate. This suggests that free-recall instructions are particularly helpful for people who perceive themselves as having good health. Perhaps these respondents see each medical-provider visit as a relatively unique event. The best way for these respondents to recall medical-provider visits may be to retrieve the personal events that happened to be associated with, or coincide with, these visits. Instructions to use a temporal retrieval order may disrupt this strategy (cf. Loftus and Fathi 1985).

Some limitations to the present approach should be noted. First, by conducting the experiment in a linkage field test designed to yield only 600 respondents, the initial sample size was relatively small. The problem of small sample size was further exacerbated by the use of respondents from the general population as opposed to preselecting respondents who had medical visits during the reference period from the rolls of HMO subscribers, as was done by Loftus et al. (1989) and by Means and Loftus (in press, studies 3–4). This resulted in the loss of respondents who reported no medical visits during the reference period and for whom we failed to obtain provider records. The statistical power to detect significant differences was thereby reduced, and may partially account for some of the marginally significant results obtained. Using a larger field study would have greatly reduced some of these problems, as would more vigorous provider follow-up. In addition, the fact that our sample was not representative of the population may have influenced some of our results. For example, self-reported health status interacted with recall condition such that free recall was most accurate for those reporting good health, but no effect-of-recall condition was found for those reporting poor health. Thus, our finding of overall superior accuracy in the free-recall condition may only be applicable for respondents who consider themselves in good health. This limitation may not be serious, however, because health surveys generally include an even higher percentage of respondents reporting good health than was found in our sample (see table 3). Fienberg and Tanur (1989) discuss other problems of this approach.

Nevertheless, the results of this first attempt to assess cognitive-recall factors using actual field-survey data are varied and interesting. Although, as might be expected, some questions remain unanswered and some tests were inconclusive, the study did yield intriguing and valuable insights into the cognitive processes of survey respondents. Perhaps the most interesting findings were that respondent gender, self-reported health status, and number of visits affected the accuracy of recall. These findings require replication on another, preferably larger sample, and, if supported, would have implications for the way in which respondents with particular characteristics should be instructed to recall information. These preliminary data suggest that if a researcher is interested in accurate codable responses, free recall is recommended. This recommendation is only valid when most respondents consider themselves to be in good health and the total number of visits is five or less. Free recall may also be the method of choice when underreporting is a concern, although backward-recall instructions would be preferred when the sample is composed primarily of males. When the number of events to be recalled is greater than five, other techniques, such as the cognitive strategies demonstrated by Means and Loftus (in press, study 4) are most effective. (These recommendations are only preliminary, and must be considered along with the limitations discussed above.) Future research might explore whether similar results are found with other autobiographical events.

Finally, our results contribute to a growing body of literature demonstrating that cognitive methodology and theories are effective in improving the quality of survey data (see Jobe and Mingay in press). Future research will determine the most appropriate applications for cognitive experiments embedded in field surveys, as well as the degree to which they are an effective bridge between the two disciplines. Such research will be greatly aided by the advent of computer-assisted face-to-face and telephone interviewing. Moreover, if different recall procedures are determined to be more accurate, depending on respondent gender and self-reported health status (as the present preliminary results suggest), or on variables identified by future research, computer-assisted interviewing will enable these differential questioning procedures to be more easily instituted as well.

References

Blair, E., and S. Burton. 1987. Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions. *Journal of Consumer Research* 14:280–88.

Blair, E.A., and G.K. Ganesh. In press. Characteristics of Intervalbased Estimates of Autobiographical Frequencies. *Applied Cognitive Psychology*.

Fienberg, S.E., and J.M. Tanur. 1989. Combining Cognitive and Statistical Approaches to Survey Design. *Science* 243:1017–22.

Hippler, H.J., N. Schwarz, and S. Sudman, eds. 1987. *Social Information Processing and Survey Methodology*. New York: Springer-Verlag.

Jabine, T.B., M.L. Straf, J.M. Tanur, and R. Tourangeau, eds. 1984. *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Washington: National Academy Press.

Jobe, J.B., and D.J. Mingay. 1989. Cognitive Research Improves Questionnaires. *American Journal of Public Health* 79:1053–55.

———. In press. Cognition and Survey Measurement: History and Overview. *Applied Cognitive Psychology*.

Kovar, M.G., and G.S. Poe. 1985. The National Health Interview Survey Design, 1973–84, and Procedures, 1975–83. *Vital and Health Statistics*, series 1, no. 18. DHHS pub. no. (PHS) 85-1320. Hyattsville, Md: National Center for Health Statistics.

Loftus, E.F. 1982. Memory and Its Distortions. In *The G. Stanley Hall Lecture Series*, ed. A.G. Kraut, 123–54. Washington: American Psychological Association.

———. 1984. Protocol Analysis of Responses to Survey Recall Questions. In *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, ed. T.B. Jabine, M.L. Straf, J.M. Tanur, and R. Tourangeau, 61–68. Washington: National Academy Press.

Loftus, E.F., and D.C. Fathi. 1985. Retrieving Multiple Autobiographical Memories. *Social Cognition* 3:280–95.

Loftus, E.F., K.D. Smith, M.R. Klinger, and J. Fiedler. 1989. *Memory and Mismemory for Health Events*. Seattle: University of Washington. (Unpublished.)

Mathiowetz, N.A., and E.P. Ward. 1987. Linking the National Medical Expenditure Survey with the National Health Interview Survey: Analysis of Field Trials. *Vital and Health Statistics*, series 2, no. 102. DHHS pub. no. (PHS) 87-1376. Hyattsville, Md: National Center for Health Statistics.

Means, B., and E.F. Loftus. In press. When Personal History Repeats Itself: Decomposing Memories for Recurring Events. *Applied Cognitive Psychology*.

Royston, P., D. Bercini, M. Sirken, and D. Mingay. 1986. Questionnaire Design Research Laboratory. *Proceedings of the Section on Survey Research Methods*, 703–7. Washington: American Statistical Association.

Schwarz, N., H.J. Hippler, and E. Noelle-Neumann. In press. A Cognitive Model of Response Order Effects in Survey Measurement. In *Context Effects in Social and Psychological Research*, ed. N. Schwarz and S. Sudman. New York: Springer-Verlag.

Smith, A.F., J.B. Jobe, and D.J. Mingay. In press. Retrieval from memory of Dietary Information. *Applied Cognitive Psychology.*

Whitten, W.B. II, and J.M. Leonard. 1981. Directed Search through Autobiographical Memory. *Memory & Cognition* 9:566–79.

Acknowledgments: Presented in part at the meeting of the American Psychological Association, New Orleans, August 1989. This research was supported in part by grants SES 84-03415 and SES 86-12320 from the National Science Foundation to the National Center for Health Statistics. E. F. Loftus received support from the National Center for Health Services Research and Health Care Technology Assessment, and from the National Institute of Mental Health. Westat, Inc. and NORC conducted the NHIS/NMES Linkage Field Test under contract with the National Center for Health Statistics and the National Center for Health Services Research and Health Care Technology Assessment. The authors thank Steven Botman for his assistance with data and several anonymous reviewers for many helpful suggestions and comments that have substantially improved this paper.

Address correspondence to: Jared B. Jobe, Ph.D., Office of Research and Methodology, National Center for Health Statistics, Room 915, 6525 Belcrest Road, Hyattsville, MD 20782.