Journal of Official Statistics
Vol. 8. No. 1, 1992, pp. 19–39

Measuring the Recall Error in Self-Reported Fishing and Hunting Activities

Adam Chu[1], *Donna Eisenhower*[1], *Michael Hay*[2], *David Morganstein*[1], *John Neter*[3] *and Joseph Waksberg*[1]

Abstract: The U.S. Fish and Wildlife Service sponsors national surveys to estimate the number of participants engaged in various wildlife-related activities, the time spent on those activities, and a wide range of related expenditures. In the past, these surveys relied on a one-year recall period about which respondents were asked about details of their activities during the preceding year. This paper summarizes the results of an experiment designed to measure the direction and magnitude of possible biases that may result from the use of a one-year recall period. The paper also examines the extent to which the bias can be reduced by using shorter recall periods.

Key words: Recall error; wildlife-related activities; episodic recall; estimation.

1. Introduction

Since 1955 the U.S. Fish and Wildlife Service has sponsored a national survey that estimates the number of participants in various wildlife-related activities, primarily fishing and hunting, the number of days spent on those activities, and a wide range of related expenditures. These surveys have been carried out at five-year intervals, with the most recent survey conducted in 1986 by the U.S. Bureau of the Census. The surveys have until now relied on a one-year recall period about which respondents are asked for details of their wildlife-related activities during the preceding calendar year. Some observers have questioned the accuracy of the estimates because of the one-year recall period.

As a result of these concerns, the U.S. Fish and Wildlife Service supported a research study to determine if a one-year recall period imparts a serious systematic bias to the survey estimates. If so, the study was to explore the extent to which a shorter recall period or other modifications to the survey's methodology could result in more accurate estimates of hunting and fishing activities. The results of the analysis were used by the U.S. Fish and Wildlife Service in making major changes for future surveys and, more broadly, add to the body of knowledge regarding respondent bias when respondents are asked to recall frequency of activities under varying lengths of the recall period.

This paper summarizes the results of a designed experiment comparing estimates of the number of fishing and hunting participants, the number of trips, days of activity,

[1] Westat, Inc., 1650 Research Boulevard, Rockville, MD 20850, U.S.A.
[2] U.S. Fish and Wildlife Service, Department of the Interior, Washington, D.C. 20240, U.S.A.
[3] University of Georgia, Athens, Georgia 30602, U.S.A.

and associated expenditures during 1988, reported with recall periods of one year, six months, three months, one month and two weeks. More detailed information is contained in the report on the project submitted by Westat to the U.S. Fish and Wildlife Service of the Department of Interior (Westat 1989).

In addition to providing information on the direction and magnitude of possible biases that may result from use of a one-year recall period, the paper also examines the degree to which such biases can be reduced by using shorter recall periods to obtain the same information. The study data permit separate assessments of bias in estimates of various types of expenditures (e.g., travel expenses, equipment purchases) and the frequency of fishing and hunting activities (e.g., number of trips made when fishing in the spring and summer, big game hunting in the fall) since these statistics may be affected differently by the length of the recall period. The effectiveness of the use of "key events" to enhance recall accuracy was also examined. Bounding procedures and memory aids such as maps and flashcards were used.

There are of course other potential sources of bias in the survey which could affect the accuracy of responses. Among these are the definition of a hunting or fishing trip particularly for extended trips and the respondent's understanding of included and excluded expenditures. This study was expressly limited, however, to the relation between the length of the recall period and recall error.

The paper contains six sections. We begin with a summary of relevant previous research on the bias associated with length of recall. Next, we provide a description of the study design, which is followed by a discussion of the analytical results. Then, we consider the sampling implications of balancing bias from recall with sampling error, a common design trade-off. Lastly, we summarize our findings.

2. Relevant Research on the Length of the Recall Period

To assist in planning our empirical study, a review was conducted of other studies of the effects of varying lengths of the recall period on the accuracy of survey data. These other studies addressed a variety of survey topics, and thus are not all directly applicable to surveys of recreational fishing and hunting; however, there was general agreement with a reasonable expectation that the longer the recall period, the less accurate the survey results.

Methodologists have examined recall errors related to the length of the recall period and have placed most of the causes of error into two major categories, namely: telescoping and recall decay. Telescoping refers to the tendency of respondents to report events in the wrong time period. In other words, respondents may report events earlier or later than they actually occur. Recall decay refers to the inability of the respondent to recall all of the relevant events occurring in the past leading to errors of omission. Telescoping is generally thought to dominate errors of recall when the reference period for the questions is of short duration and the events to be remembered are infrequent and large. Recall decay is more likely to have a major effect when the reference period is of long duration and the events to be remembered are frequent and routine.

There does not seem to be much information on whether the effects of telescoping and recall decay are similar when the respondent is asked to use episodic recall or to estimate the frequency of a given event. In episodic recall, the respondents are either instructed, or decide on their own, to recall

each event separately when reporting a total frequency of occurrence of an activity. Estimation is when the respondent provides a total number by using an estimation strategy such as "I do this event approximately once a week so I must do it about 52 times in the course of the year."

Our study suggests that under circumstances like the ones in attempting to estimate fishing and limiting experiences, other factors have a greater effect on the direction and number of recall errors than telescoping or decay. Furthermore, the length of the recall period appears to influence the size of the recall error caused by these factors. Most, although not all, studies have reported underestimates as the length of the recall period increases. Our study shows the opposite effect, and what is suggested is that some factor related to social desirability, prestige, or possibly one's memory of pleasant events is involved when estimation is in use. When respondents resort to the use of estimation strategies in providing numerical answers then the direction of the error in providing an educated guess is related to how desirable or prestigious it is to provide a relatively high or low number of occurrences of the event. This pattern is supported by the recall literature.

2.1. Findings from recall studies of recreation and recreational fishing and hunting

The studies related to recall errors in estimates of recreational participation do not have consistent results. Chase and Harada (1984) reported on the accuracy of estimates of 800 respondents' self-reports of swimming activity as checked against pool records for the 1981 swimming season. The results indicated that recall error produced an overestimation of actual participation. The respondents with the largest reports of participation typically had the largest amount of error in their estimates. Furthermore, the more important or salient swimming was to the respondent, the greater the percentage error. This suggests that some factor related to social desirability, prestige bias, or other psychological factors influences the direction of the bias. Finally, this study found digit bias in the self-reports with almost all estimates of frequency of participation being numbers ending in 0 or 5.

Similarly, Chase and Godbey (1983) conducted two studies comparing club records with self-reported activity at a tennis club and at a swimming pool as derived by mailed questionnaires. Both studies revealed that more than 75% of respondents overestimated their frequency of participation, with at least 43% in both studies overestimating their frequency of participation by more than 100%. As in the previous studies, the data suggest that overestimates are greater for those with higher skill levels and those who place greater importance on the activity.

Gems, Ghosh, and Hitlin (1982), on the other hand, found that a two-month recall period produced underestimates when compared to a collection every two weeks. They conducted an experiment to investigate the optimum recall period for anglers to accurately report their fishing activity. The study was devised for the National Marine Fisheries Service and compared reports from people contacted by telephone at the end of two months to those contacted every two weeks for a two-month period using bounded recall procedures. The results of this study indicated that the two-month recall period using a household telephone survey produced underestimates of true fishing activity. The true activity level was defined as the estimates from the data collected every two weeks. In addition, significantly more anglers reported *no* activity using a

two-month recall period than a two-week interval, and avid anglers forgot trips the most.

An earlier methodological study of 138 anglers, also conducted for the National Marine Fisheries Service (Ghosh 1978), found an overestimate with longer recall. However, the methodology for this earlier study was different from the one conducted in 1982. In this study, the anglers were first asked to estimate the number of recreational fishing trips they took in the last 30 days. They were also asked to list the dates of those trips, which were already known to the interviewer, as far back as they could remember beginning with the most recent. An analysis of the comparative data reveals that in giving estimates of the total number of trips in a 30-day time period, anglers' magnitude of telescoping error (overestimates) is probably greater than omission error (underestimates). Both the 1978 and 1982 studies have relatively short lengths of recall periods suggesting that respondents were more likely to use episodic recall. For episodic recall, telescoping and recall decay are likely to influence the accuracy of the results.

2.2. *Findings from other recall studies*

A substantial methodological literature has developed since the 1950s, which has examined the effects of varying the length of the recall period on the accuracy of survey results. Many of the studies compared respondent reports to actual records. Many studies were conducted in subject areas that lent themselves to record comparisons. A large number of these studies fall into one of three areas, namely: studies of consumer expenditures and earnings; studies of hospital data and medical records; and studies of crime and accidents.

Some of the earliest studies on the accuracy of recall involved consumer expenditures (Ferber and Birnbaum 1979; Jaegar and Pennock 1962; Metz 1956; Neter and Waksberg 1964; Politz 1958; Sudman 1964; Sudman and Ferber 1970). Previous recall studies were also done with respondent information which could be verified with hospital and medical records: (Cannell 1965; Cohen, Erickson, and Powell 1984; Gray 1955; Means, Mingay, Nigan, and Zarrow 1988; the National Center for Health Statistics 1962, 1965a, b, 1967; and Woolsey 1953). There have also been a number of comparisons of respondents' reports of crimes, injuries, and motor vehicle accidents with police records (Cash and Moss 1969; Murphy and Cowan 1976; Tracy and Fox 1981; Waksberg and Valliant 1978; Woltman, Bushery, and Carstensen 1975; and Wyner 1980). The majority of these studies verified the numerical answers produced by respondents with actual records. This empirical literature, in general, supported the relationship that longer recall periods lead to less accurate results. During a longer recall period, a larger number of events may occur and, thus, must be recalled. However, too short a recall period may also invite errors related to telescoping. Most of these studies relied on episodic recall rather than estimation so that telescoping and recall decay are germane. The majority suggest that longer recall periods usually produce greater error in the direction of underreporting. As such, there is underreporting of minor occurrences or events and telescoping for more important events with recall decay acting as a stronger source of error for the majority of studies.

As mentioned earlier, other factors appear to influence the direction of error in addition to telescoping and recall decay. Perhaps social desirability or prestige bias may also be influencing the direction of the error when the recall period is long and all

the relevant events cannot be adequately recalled. Under these circumstances, underreporting tends to occur for these less desirable events such as expenditures, hospitalizations, illnesses, and arrests.

3. Study Design

The major objective of the design was to compare respondents' abilities to recall trip characteristics for varying recall lengths. The recall lengths examined ranged from two weeks to one year. The use of key events in facilitating recall for long recall periods was also studied. These objectives were met through data collection from six respondent panels selected for the study by random digit dialing procedures in the study areas.

The panels used in this study were obtained from a probability sample drawn from East Texas and East Central Wisconsin. These areas were selected because they have a mix of different kinds of hunting and fishing, while the average of their utilization is very similar to the national norm.

Households were called at random in January 1988 and the respondents were asked if anyone 16 years of age or older hunted or fished during 1987. Households with active anglers and hunters were asked to participate in this study during 1988 and were stratified by the level of fishing and hunting activity they reported for 1987. The "active" households in each stratum were randomly assigned to one of the six panels, each having a different recall procedure.

The study sample consisted of 2,866 "active" households identified in the screening of 8,752 households. After being screened into the study on the basis of fishing/hunting activity in 1987, each household was subsequently contacted by an interviewer. Table 1 summarizes the six experimental groups (panels) used in the design. Panels A through D were screened by telephone by a field interviewer for their activity during the stated period. If a respondent was active during the period, a field interviewer administered the main questionnaire in person. Due to budgetary limitations, panels E and F (the monthly panels) were administered a questionnaire by a telephone interviewer in a central telephone facility.

To provide a two-week recall condition without incurring the additional cost of such frequent contact, panels E and F were interviewed for overlapping time periods. Panel E respondents were first asked for details about hunting and fishing trips which took place from the 1st to the 15th of the month and then from the 16th to the 30/31st of the month. Panel F respondents were asked about these activities for the 16th of the same month to the 30/31st and then the 1st to the 15th of the next month. This overlapping design provided statistical comparisons of a two-week recall period (actually semi-monthly) with a monthly recall period without the added costs and respondent burden required by a two-week contact period. Whereas the respondents in panels A through D were interviewed regarding an entire year's experience, respondents in panels E and F were interviewed for only a nine-month period because of unforeseen delays in preparing materials for the field survey.

We do not believe that the interview modality, telephone vs. in person, had a serious effect on the study results. As shown in the next section, the trends found with increasing length of recall apply to all of the panels whether contacted in person or by telephone. Since the results are consistent across a range of outcomes, they do not indicate that modality is a factor.

As mentioned in Section 2, most of the research on recall length support the reasonable expectation that shorter recall periods result in more accurate results. Our results

Table 1. Panel organization, sample size, and interview frequency

Panel	Interview method	Description	Target sample size (households)
A	Field – in person	Annual recall using nonkey event	500
B	Field – in person	Annual recall using key event	500
C	Field – in person	Six-month recall	500
D	Field – in person	Three-month recall	600
E	Telephone	Monthly – 1st to 30/31st of month	350
F	Telephone	Monthly – 16th of month to 15th of following month	350

indicate a clear pattern of increasing number of events per unit time recalled when using longer recall periods. In addition, the differences increase with greater recall length. Thus, if previous research conclusions are correct, the relatively insignificant differences between two-week and monthly recall periods as contrasted with the more noticeable increases between quarterly, semi-annual and annual recall suggests that the shorter periods we tested would provide the most accurate responses.

In summary, the study design established an experimental base to assess differences in accuracy of estimates of trips, days, and expenditures for fishing and hunting activities using recall periods of differing lengths.

4. Analysis of Results

4.1. Combining selected panels

In most of the tables, the two annual panels A and B, and the two monthly panels E and F have been combined for analysis purposes. In the original experimental design specifications, panels A and B were established in order to determine the extent to which the use of "key event" to define reference periods within the year could improve the accuracy of recall. However, the results from the two annual panels are virtually identical with respect to the key statistics mentioned above, and therefore the two panels were pooled in the subsequent analyses. Similarly, the two monthly panels E and F, which were originally designed to estimate the effects of a two-week recall, were combined because the results for two-week and monthly recalls were similar.

4.2. Factor effects analyzed from estimated annual averages

Table 2 summarizes the weighted annual averages for fishing-related statistics for the original (unpooled) panels A through D, and the corresponding nine-month averages for panels D9, E, and F. Bounded data for the monthly panels were obtained only for the last nine months of 1988. The notation "D9" is used to refer to the results from the last nine months of the quarterly panel, D, and does not represent a separate and distinct panel. The purpose of the weighting was to compensate for respondent losses which occurred during the survey period and which produced small differences in the mix of avid and nonavid sportspersons allocated to the various panels.

Also shown are the annualized results for panels E and F denoted in the table as E[ann.] and F[ann.], respectively. "Annualized" results for panels E and F are presented to permit direct comparison with panels A

Table 2. Weighted averages and standard errors for fishing

Site	Panel	Proportion of persons who reported taking at least one fishing trip		Average number of fishing trips		Average number of fishing days	
		Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Both	A	0.75	0.02	21.90	1.58	26.95	1.71
	B	0.75	0.02	20.26	1.39	26.65	1.59
	C	0.69	0.02	17.97	1.29	21.30	1.34
	D	0.62	0.02	16.58	1.00	20.05	1.06
	D9	0.60	0.02	13.52	0.84	16.74	0.89
	E	0.66	0.02	8.99	0.77	11.49	0.85
	F	0.68	0.02	8.84	0.77	11.80	0.85
	E[ann.]	0.68	0.02	11.02	0.94	13.76	1.02
	F[ann.]	0.70	0.02	10.84	0.94	14.13	1.02

Site	Panel	Average fishing trip-related costs		Average fishing equipment purchases	
		Estimate	Standard error	Estimate	Standard error
Both	A	442.68	30.69	117.25	8.84
	B	534.03	66.15	132.66	13.71
	C	416.74	30.21	134.25	12.58
	D	343.79	23.29	98.80	8.03
	D9	258.46	19.21	69.45	6.37
	E	173.32	13.88	44.54	5.90
	F	196.68	17.87	59.36	14.51
	E[ann.]	230.54	18.46	63.37	8.40
	F[ann.]	261.61	23.77	84.44	20.64

through D. These annualized averages were obtained by multiplying the averages from panels E and F by the ratio of the corresponding average for panel D to the average from panel D9.

It can be seen that key events do not appear to make a difference. The results for the with-key event (panel B) and the panel not using key events (panel A) are generally similar for each of the five statistics considered. None of the differences between the two panels was statistically significant. Similarly, for the fishing-related statistics, no statistically significant differences between the two monthly panels E and F were found.

A fairly clear pattern of recall period effects is evident for fishing. The number of trips, days, and costs for the annual panels are all higher than those for semi-annual recall which in turn are higher than those for quarterly recall, which are higher than with monthly recall. Differences between the annual and quarterly panels are highly significant for the estimated proportion of persons who fished, average number of days of fishing, and average trip-related costs

Table 3. Weighted averages and standard errors for hunting

Site	Panel	Proportion of persons who reported taking at least one hunting trip		Average number of hunting trips		Average number of hunting days	
		Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Both	A	0.52	0.02	14.59	1.06	19.55	1.16
	B	0.48	0.02	16.37	1.68	20.81	1.69
	C	0.40	0.02	12.53	0.93	16.75	1.02
	D	0.41	0.01	13.17	0.85	17.29	0.92
	D9	0.39	0.01	12.09	0.81	16.08	0.88
	E	0.42	0.02	10.65	0.98	14.66	1.02
	F	0.39	0.02	9.91	0.85	14.25	0.98
	E[ann.]	0.44	0.02	11.60	1.07	15.76	1.10
	F[ann.]	0.41	0.02	10.80	0.93	15.32	1.05

Site	Panel	Average hunting trip-related costs		Average hunting equipment purchases	
		Estimate	Standard error	Estimate	Standard error
Both	A	274.26	25.04	237.35	30.45
	B	235.27	20.56	203.51	16.66
	C	267.06	30.58	276.76	40.12
	D	236.70	20.33	235.13	31.61
	D9	203.54	15.91	154.85	25.02
	E	194.63	18.45	124.74	15.58
	F	193.26	18.15	112.85	17.33
	E[ann.]	226.34	21.46	189.41	23.66
	F[ann.]	224.75	21.11	171.36	26.31

(p-value $< .01$ for any given comparison). Moreover, the observed difference between semi-annual and quarterly is also significant for the proportion of persons who fished and average fishing equipment purchases (p-value $< .01$). For trip-related costs, the difference between semi-annual and quarterly is moderately significant (p-value $\approx .02$); however, the direction of the differences is consistent with the overall pattern. (These patterns appear to hold when the data are disaggregated into individual sites as well.) Similarly, the differences between quarterly and monthly data are almost all significant and in a direction consistent with a hypothesis that increased length of recall period results in higher estimates.

The corresponding estimates for hunting are summarized in Table 3. The key-event panels yielded results very similar to the panels without key events, and the two monthly panels were also about the same. In general, a key event did not produce significant differences (at the 0.05 level). Similarly, differences between panels E and F were not significant.

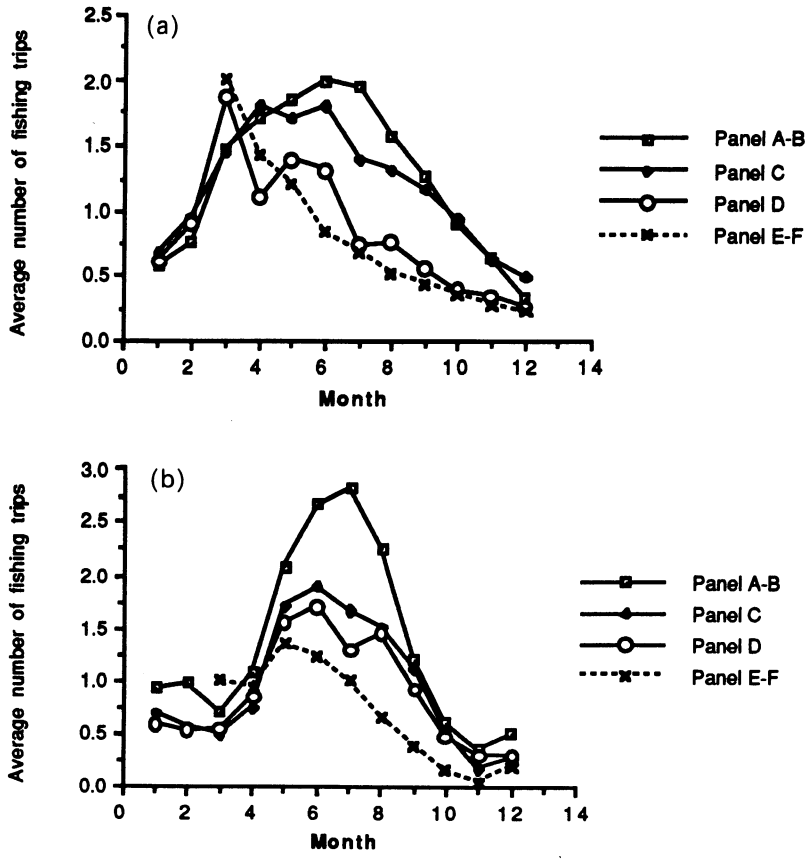

Fig. 1. (a) Average number of fishing trips reported by anglers in Texas. (b) Average number of fishing trips reported by anglers in Wisconsin

The recall effect patterns shown in Table 3 for the hunting statistics are generally similar to those observed earlier for fishing. However, there is one important difference: the results reported over varying recall lengths appear to be much closer for hunting than for fishing. This suggests that, for hunting, data from semi-annual or quarterly panels may be subject to acceptably small recall bias.

Because hunting occurs primarily in the fall and early winter, the fact that the proportion of persons who reported taking a hunting trip is significantly higher in panel A–B than in the other panels is puzzling because of the length of recall is effectively the same for panels A–B, C, and D.

4.3. Number of fishing trips by month

Figures 1a and 1b show that the number of trips reported by respondents in both sites tends to be much higher with annual recall (panel A–B) than for either quarterly (panel D) or monthly recall (panel E–F), especially in the spring and summer months. The fact that similar differences among panels occur in almost all months makes it unlikely that telescoping is a major cause of the pattern.

4.4. Number of hunting trips by month

Differences in the monthly averages for all types of hunting trips were relatively small

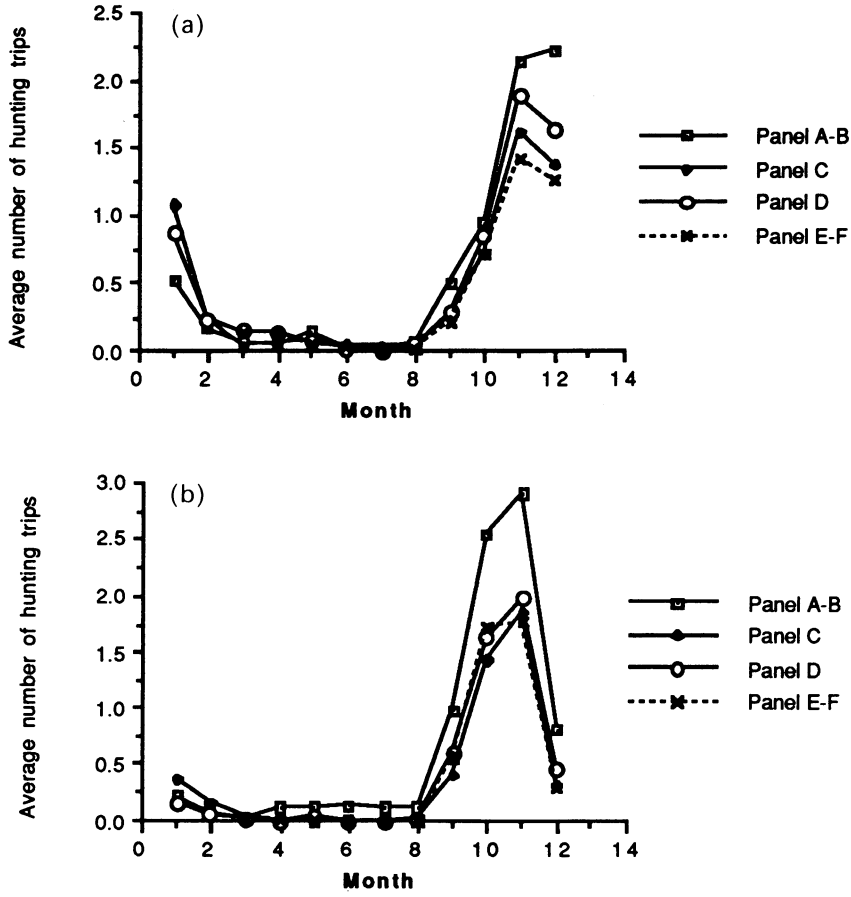

Fig. 2. (a) Average number of hunting trips reported by hunters in Texas. (b) Average number of hunting trips reported by hunters in Wisconsin

as compared with those observed for fishing. The largest differences among the various panels were observed for the late autumn/early winter months; i.e., the months that coincide with the hunting season (see Figures 2a and 2b). As was the case for fishing, respondents in panel A–B in both sites reported significantly more trips than respondents in panel C or D. However, unlike fishing, differences among the non-annual panels (panels C, D, and E–F) were not statistically significant.

4.5. Effects related to overstatements

As noted above it is unlikely that telescoping is the major reason for these differences; other forms of recall bias such as a tendency to exaggerate or other reasons for overstating the number of trips may have a greater effect on the apparent overestimates.

Figures 4a and 4b show the monthly "overstatement" in the number of fishing trips for panels A–B, C, and D as compared with panel E–F. We use the term "overstatement" based upon the previous research cited earlier which indicated that the most accurate recall occurred with the shorter recall periods even though we have no independent assessment of recall accuracy. Indicated in the figures (above the horizontal axis) are the points in time in which the interviews were conducted for the various

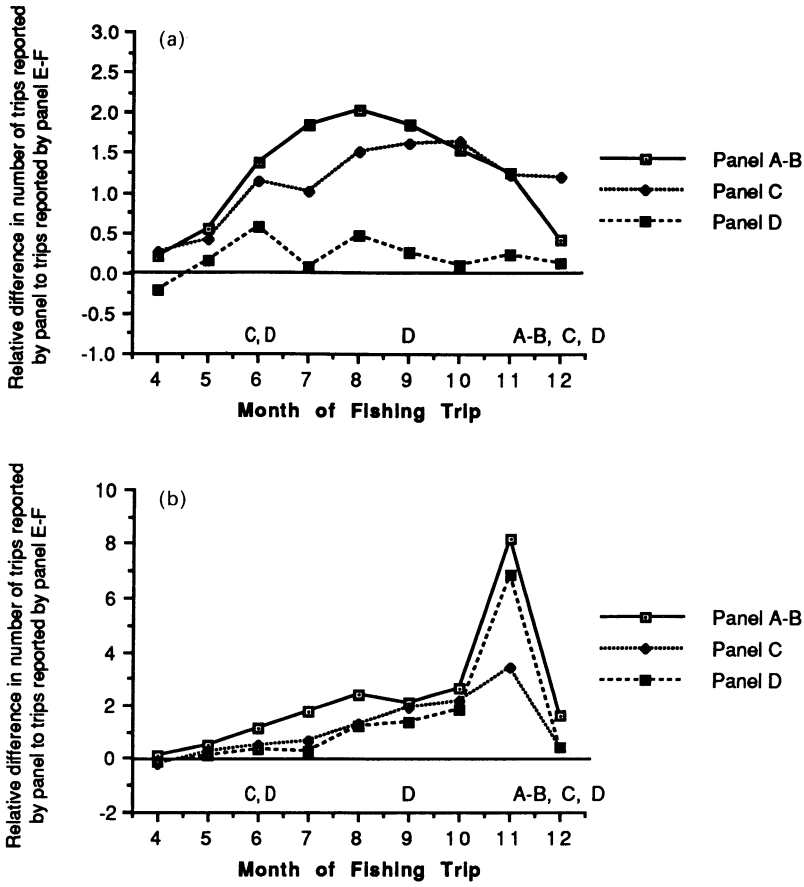

Fig. 3. (a) Overstatement of the average number of fishing trips for panels A–B, C, and D – Texas. (b) Overstatement of the average number of fishing trips for panels A–B, C, and D – Wisconsin

panels. If we take the estimates of numbers of trips provided by the monthly panel E–F to be close to the true values, then for a given month, the overstatement can be expressed as a relative difference in the number of trips reported in that month by a particular panel compared to the corresponding number reported by members of panel E–F. A value of 0 indicates no difference, while a value of 1 indicates that the panel reported 100% more activity than the comparison panel E–F.

It appears from these figures that the degree of overstatement in the average number of fishing trips is highest for the months corresponding to periods of greatest activity, namely the spring–summer months. An exception to this is the large relative difference observed for Wisconsin in November (see Figure 3b). This large relative difference is due to the fact that the number of trips reported to have been taken in November by members of panel E–F was close to zero. Excluding this unusual value, the relative differences for both Texas and Wisconsin are generally higher in the months of June to September. Thus, while there is a tendency by respondents in the annual and semi-annual panels to overreport the number of trips for all time periods, there also appears

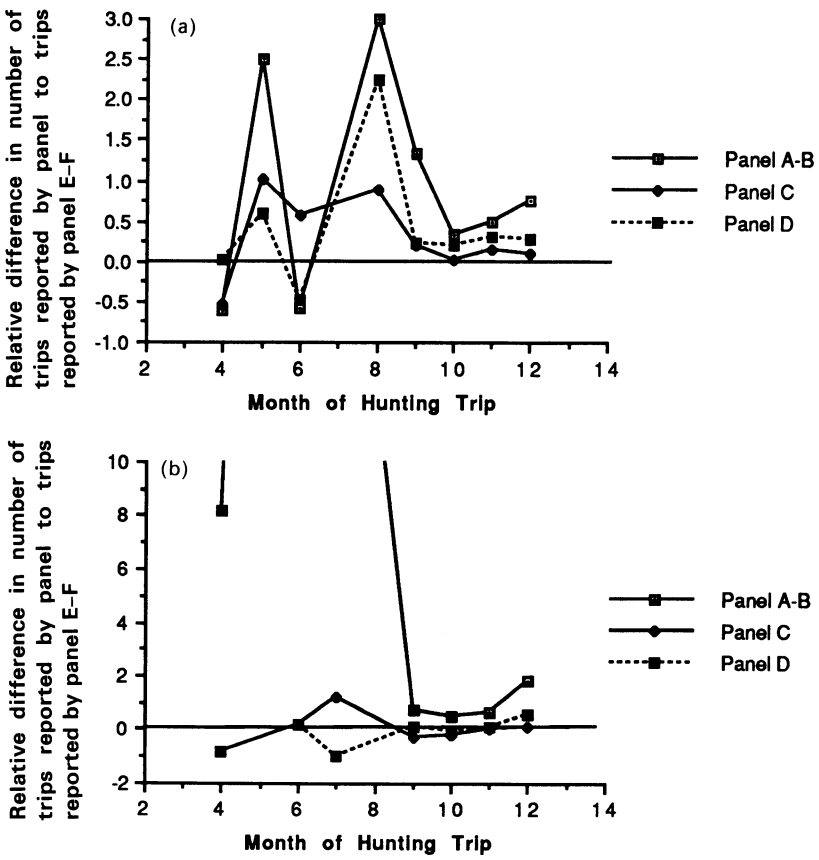

Fig. 4. (a) Overstatement of the average number of hunting trips for panels A–B, C, and D – Texas. (b) Overstatement of the average number of hunting trips for panels A–B, C, and D – Wisconsin

to be a tendency to report extra trips in those months in which one might expect the highest levels of activity. It is during these time periods that a respondent is more likely to use estimation strategies in providing numerical answers rather than episodic recall. Anglers may provide overestimates because their pleasant memories exaggerate the number of events, or it is more desirable or prestigious to have high rather than low estimates.

For hunting, where events tend to be less routine and less frequent than for fishing, the corresponding relative differences are highly erratic (see Figures 4a and 4b. Some points for panel A-B are off of Figure 4b). Again, this is due to the extremely small number of trips reported by members of panel E–F for the months prior to September. However, for the months of September to December when considerable hunting takes place, the relative differences are relatively stable and constant for panels C and D. For panel A–B, the relative differences for these months are consistently higher than for panels C and D. Also for panel A–B, the relative difference in hunting trips for December is higher than for either of the preceding two months. This may reflect that respondents included trips they actually took in the following January in December since January is a heavy hunting season

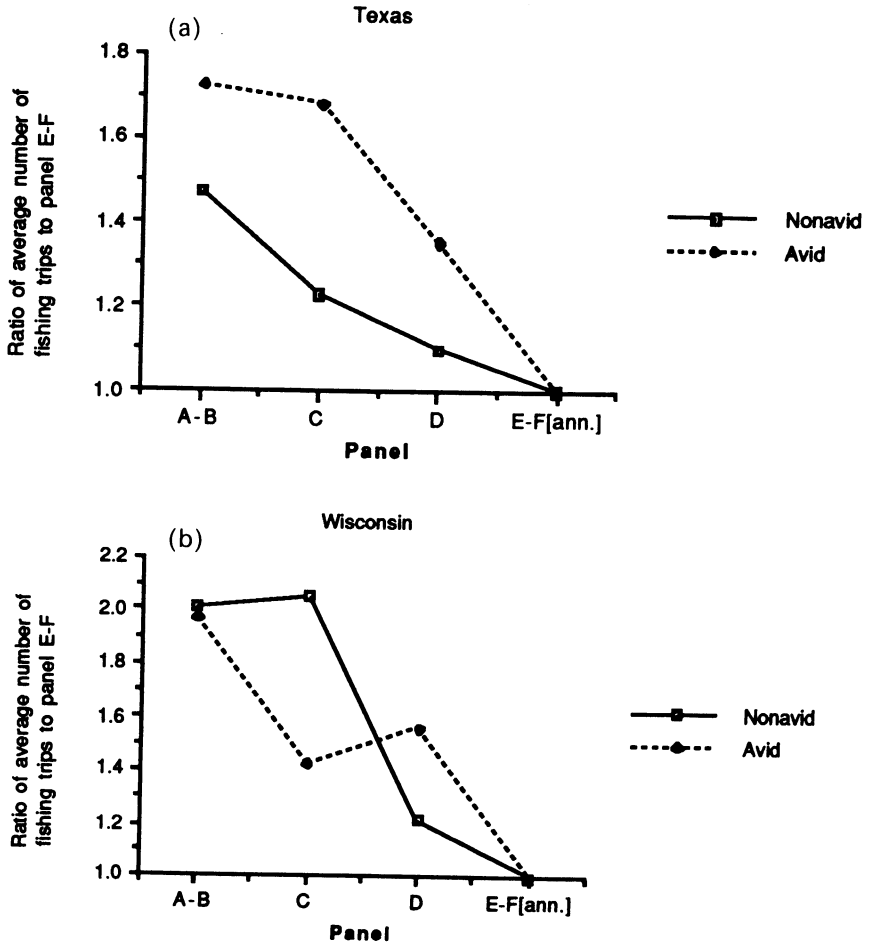

Fig. 5. (a) Ratio of average number of fishing trips reported by persons in Texas to average in panel E–F, by panel and avidity. (b) Ratio of average number of fishing trips reported by persons in Wisconsin to average in panel E–F, by panel and avidity

and the interview could have taken place any time between January 1 and mid-February.

4.6. Comparison of results by avidity

To examine the extent to which avidity affected recall, separate estimates of the fishing and hunting statistics presented earlier were prepared for "nonavid" persons (i.e., persons in households reporting 10 or fewer days of fishing or hunting activities during the prior year) and "avid" persons (i.e., persons in households reporting more than 10 days of fishing or hunting activities during the prior year). The classification of persons into the avidity strata was made on the basis of the initial screening survey, not the information obtained in the interview. The results for avid and nonavid sportspersons are summarized graphically in several figures. Note that in these figures, the panel estimates are plotted on a relative basis, i.e., the plotted points represent the ratio of the panel estimate to the corresponding estimate from panel E–F.

Avid and nonavid sportspersons show

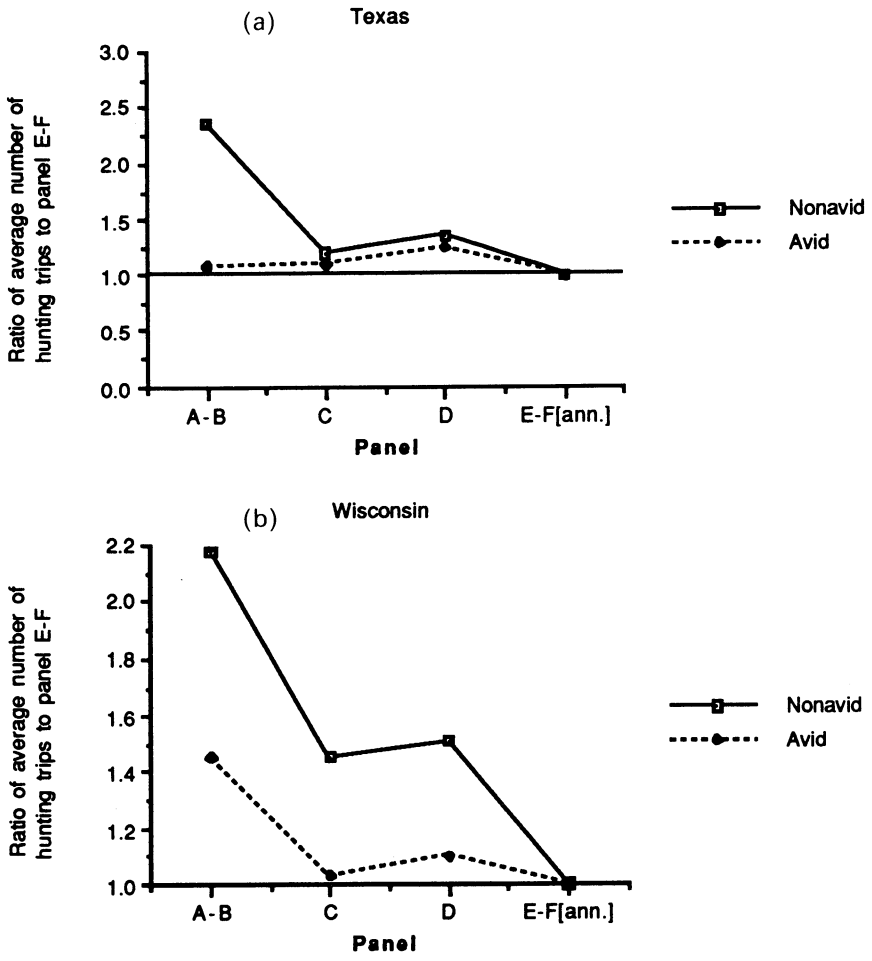

Fig. 6. (a) Ratio of average number of hunting trips reported by persons in Texas to average in panel E–F, by panel and avidity. (b) Ratio of average number of hunting trips reported by persons in Wisconsin to average in panel E–F, by panel and avidity

similar patterns in panel effects, with avid anglers revealing greater overstatements than those who are less avid. In particular, for both avid and nonavid anglers, there is significant overreporting of fishing trips when annual recall is used (see Figures 5a and 5b). However, the magnitude of the overstatement is generally smaller for the nonavid anglers than it is for avid anglers.

The corresponding results for the average number of hunting trips are shown in Figures 6a and 6b. In Texas, the average number of trips reported by nonavid members of panel A–B is significantly higher than that reported by nonavid members of either panel C (p-value $< .004$) or panel D (p-value $< .01$). Among the avid hunters in Texas, these differences were not significant. In contrast, the results for Wisconsin indicate that the average number of trips reported by avid members of panel A–B is significantly higher than that reported by avid members of either panel C (p-value $< .004$) of panel D (p-value $< .01$), while no significant differences were observed among the nonavid hunters.

5. Optimal Design

5.1. Balancing mean square error and cost

Looking at bias alone is not sufficient for decisions on sample design. When several alternative procedures are available for data collection in a survey, a prime consideration in choosing among them is a comparison of the mean square errors (MSE) of key statistics for a given total budget. We follow the usual practice of defining the MSE of a statistic as the sum of the variance and square of bias. The bias differs among the statistics collected as well as among procedures. The variance depends on the sample size and sample design. In practice, the sample designs for all procedures contemplating face-to-face interviews are the same. RDD samples would have lower design effects.

In the present context, the bias is measured by the error in the estimate due to recall loss or arising from other procedural effects. All surveys, including those on fishing and hunting, have other sources of biases as well but it seems reasonable to assume they are independent of the periods of recall and will not affect comparisons among procedures. They can then be ignored in choosing among alternative recall lengths and contact strategies.

We estimated the MSE for the statistics that most analysts would consider of prime importance. We estimate these statistics for both the total population and a subgroup which amounts to 10% of the total.

In almost all multi-purpose surveys comparisons of MSEs for alternative procedures usually indicate that no single procedure is best for all statistics. For fishing and hunting surveys, lengths of recall period that are best for analyses of anglers do not produce minimum MSEs for all statistics on hunters. Furthermore, even for a single statistic, the best procedure for the total population is not always the best for various subgroups. Although these results are fairly common in survey research, we were surprised at how different the conclusions were depending on which statistic and subgroups were being examined. Consequently the priorities attached to various classes of statistics become a crucial issue.

We commented earlier that there are other sources of reporting errors, most of which are approximately identical among the procedures. To the extent they differ, they should be taken into account. For example, if one procedure has only slightly higher mean square errors for most items than another procedure, but it provides a much higher response rate, or it appears to be significantly better for one or two important items, it may be preferable to use that procedure even though it has a slightly higher MSE.

5.2. Estimates of bias

The analyses of MSE were restricted in several ways.

1. The statistics examined were estimates of the proportion of persons who went fishing and hunting, their average number of days spent in these activities, and their average expenditures related to hunting and fishing. The agency sponsoring the study indicated that these were the key items to study.
2. We examined these items for a "typical" state. The national surveys use very large sample sizes so that they can produce data for each of the 50 states. The total United States variances are thus very low and it is clear that minimizing the MSE for national statistics requires minimizing the biases, that is, using a monthly panel. However, since state data are crucial goals of the survey, it is necessary

Table 4. Estimates of squared bias and variance for an average site state

Model parameter	Proportion of persons		Average days		Average expenditures	
	Fishing	Hunting	Fishing	Hunting	Fishing	Hunting
Variance						
Area sample	.000144	.000064	2.40	5.81	5625	10000
RDD sample	.000111	.000049	1.85	4.47	4331	7700
Square of bias						
Annual	.001156	.001024	132.48	17.47	51040	977
Semiannual	.000256	.000225	27.35	3.61	10545	202
Four-month	.000100	.000081	9.86	1.30	3796	73
Quarterly	.000049	.000036	4.37	.58	1688	32
Monthly	.000000	.000000	.00	.00	0	0

to examine the effect of recall procedures on state statistics.

The tabulated data show a somewhat irregular progression in moving from larger to shorter recall periods. It is very unlikely that such irregularities reflect the effect of memory for the various time periods. We believe they are due to sampling error or other survey problems. We have consequently smoothed the series of estimates by fitting a line, separately for each statistic, to the four observations. Note that the bias for a four-month recall given in Table 4 was estimated from this fitted line.

5.3. *Estimates of variance*

Since the primary goal of the research was to provide guidance on plans for a national survey, the variances and design effects that should be included in the MSE are those that would come from a national survey. The experimental study variances are not appropriate for this analysis, although they are relevant to analyses of whether differences among the various experimental procedures are statistically significant. We estimated variances for an "average" state by averaging relvariances in the 1985 national survey for a number of geographically diverse states, and converting these to average variances.

5.4. *Mean square errors*

Table 4 summarizes estimates of variances and bias squared for an average state for the key statistics. The table presents estimates of variance for both area samples and RDD survey. Biases are shown for four recall periods including a four-month recall period. The bias for the four-month recall was estimated indirectly by fitting a straight line to the results for the recall periods tested in the experiment. We assume monthly data do not have any recall biases.

Tables 5a and 5b combine the variances and biases into mean square errors. MSEs are shown for both estimates relating to the total population and to subsets consisting of 10% of the population. The subsets could include age groups; different types of fishing and hunting, for example, salt water vs. fresh water fishing, or big game vs. small game hunting, fishing in specific regions of the state, etc. The calculations assume the biases are identical in the subsets and that the variances are inversely related to sample sizes.

Tables 5a and 5b show that the improvement in the MSE in going from annual, to

Table 5a. Mean square errors of alternative recall and interview procedures for state data for statistics on 100% subdomain

Procedure	Average expenditures		Average days	
	Fishing	Hunting	Fishing	Hunting
Annual	56,665	10,977	134.88	23.28
Semi-annual	16,170	10,202	29.75	9.42
Four-month	9,421	10,073	12.26	7.11
Quarterly	7,313	10,032	6.77	6.39
Monthly w/area	5,625	10,000	2.40	5.81
Monthly w/RDD	4,331	7,700	1.85	4.47

Table 5b. Mean square errors of alternative recall and interview procedures for state data for statistics on 10% subdomain

Procedure	Average expenditures		Average days	
	Fishing	Hunting	Fishing	Hunting
Annual	107,290	100,977	156.51	75.55
Semi-annual	66,795	100,202	51.38	61.69
Four-month	60,046	100,073	33.89	59.38
Quarterly	57,938	100,032	28.40	58.66
Monthly w/area	56,250	100,000	24.03	58.08
Monthly w/RDD	43,310	77,000	18.50	44.72

semi-annual, and progressively down to monthly data collection varies greatly among the items, and on whether one examines statistics for the total or for one of the subsets. The different effects reflect the varying relationships between bias squared and variance. For estimates of 100% of anglers and hunters, the bias dominates and substantial improvements occur with shorter recall periods. For small subsets, the variance dominates, and the resulting MSE is about the same for the shorter recall periods.

For statistics on average days, there is a striking reduction in MSE for average number of days spent fishing. Most of the reduction occurs in going from annual to quarterly. Monthly data show a continued improvement for estimates of all anglers but only a minor gain for a 10% subset. The improvements in having shorter recall periods are much smaller for average days spent hunting. There is a sharp reduction in the mean square error for the total population in going from annual to semi-annual, with only minor gains for subsequent changes in recall lengths. The improvements for subsets are even more modest.

Average expenditures for fishing show about the same patterns as average days; there is a reduction in the MSEs going from annual to semi-annual, with more modest gains after that. The average expenditure on hunting, however, has a different pattern than the other items. The variance so dominates the MSE that the recall period has hardly any effect. The MSE is virtually the same for all recall periods.

We note that since it is more expensive to carry out multiple interviews throughout the year than to conduct a single annual survey, the costs of the alternative recall procedures need to be considered in addi-

Table 6. Mean square errors of alternative recall and interview procedures holding interviewing costs constant (area sample)

Procedure	Average expenditures		Average days	
	Fishing	Hunting	Fishing	Hunting
Annual	56,665	10,977	135	23
Semi-annual	21,795	20,202	32	15
Four-month	20,671	30,073	17	19
Quarterly	24,188	40,032	14	24
Monthly	67,500	120,000	29	70

tion to MSE in choosing the most efficient procedure. We have not estimated the costs of various procedures; it was outside the scope of the study. However, it is obvious that short recall periods that require more frequent interviews will cost considerably more than longer recall periods. Moreover, assuming that the unit costs for interviewing are about the same for the different recall periods (i.e., interviewing one household four times a year costs about the same as interviewing four households once a year), it is possible to obtain a rough idea of the resulting MSEs when interviewing costs are held constant. Table 6 summarizes the results of such an analysis for selected statistics. It is clear that when cost is taken into account an annual recall is the most efficient method for obtaining data on hunting expenditures. For fishing expenditures it is likely to be between quarterly and semi-annual. The optimum period for average fishing days and average hunting days appears to be close to quarterly and semi-annual, respectively.

6. Implications for Survey Design

There are a number of results that stand out as having implications for future surveys, not only on fishing and hunting characteristics, but for other studies involving similar issues of memory and recall.

- There is no single procedure that is best for all statistics. Whatever method is settled on is necessarily a compromise. As a result, priorities need to be kept in mind in deciding on a survey design.
- One year recall produces substantial overestimates of most hunting and fishing characteristics.
- In general, the size of the bias diminishes with shorter recall periods, but at least for fishing where there are frequent, routine events there seems to be a significant overstatement even with the use of semi-annual recall.
- Although hunting followed the same general pattern as fishing, the biases were much smaller, probably due to the different seasonal patterns and the fact that hunting tends to be a less frequent routine activity.
- When mean square errors at a constant cost are taken into account, it appears as if quarterly or three times a year data collection is reasonably close to an optimum design for hunting items. For fishing, somewhat smaller recall periods are probably preferable.
- For fishing, the relative amount of overstatement for long-term recall is greatest during the peak months of fishing activity. The pattern was not as clear for hunting, probably because of the strong seasonal pattern and again the nature of the hunting event as bigger and less routine.

- Using key-events to enhance memory did not improve annual recall. This may be because the respondents were asked to estimate activities during given time periods rather than to itemize individual events such as is done in episodic recall.
- More detailed tabulations than could be presented here provide further evidence that the more frequent activities are remembered better than the more common ones, as has already been noted in the case of hunting versus fishing. Thus reporting was better on big-game than small-game hunting, and in Wisconsin on salt-water than fresh-water fishing. The study area in Texas consisted of mostly coastal counties, and salt-water fishing was closer to a routine recreational activity in Texas than in Wisconsin.
- The tendency to overstate fishing activities with the long recall periods appears to hold for both avid and nonavid anglers; for hunting activities, the overstatement occurs mainly among avid hunters.
- For most items, recall bias dominates the total MSE. However, variance is the more important component for hunting expenditures. The length of the recall period with a fixed survey budget should depend on the priorities for expenditures as compared to other items.

An important difference between the outcome of our study and those of most other recall studies conducted in the past 20 years is in the direction of recall bias. In the majority of earlier studies, researchers found that longer recall periods resulted in underreporting of events. However, our study consistently found evidence of an overreporting bias with longer recall periods. There could be several reasons for this difference. Possibly the most important reason is that in many of the other recall studies, respondents were asked to report on individual events, whereas in the present study, respondents were asked to summarize their experiences (e.g., by reporting the number of trips or days of fishing or hunting) rather than to itemize each event. This difference might partially explain the exaggerations in the respondents' recollections of fishing and hunting activities. In addition, subtle tendencies of respondents to overestimate activities that they enjoyed and were important to them may have contributed to the overreporting. Most of the studies in the literature involved unpleasant or undesirable activities such as hospitalization, crimes, arrests, or consumer expenditures, whereas this study addressed a pleasurable or desirable activity. Participants in focus group meetings prior to the start of the survey indicated that they expected errors in reporting to be in the direction of overestimates.

The literature on recall bias in surveys of recreational activity provides mixed results. For example, studies comparing the frequency of tennis and swimming events as reported by participants with the actual figures based on club records found recall bias in the direction of overreporting. Similarly, one of two studies on fishing found an overreporting bias in estimates of the number of recreational fishing trips. In the other study, underreporting was observed for the longer recall period. However, the two studies had very different methodologies, which could have contributed to the difference in results. Both of these earlier studies differed from the present study in that they used one- and two-month reporting periods during the autumn months and required reporting of individual events instead of aggregates of all events in a given time period.

The results of this study have been incorporated in a major redesign of the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR). The next

survey will cover 1991 and will be done with a four-month recall period and bounded interviews. That seemed to be a reasonable compromise to an optimal recall period length, taking into account the reduction in MSE, the additional costs of repeated interviews, and the logistical demands a shorter recall period would put on U.S. Census Bureau resources. The 1991 survey will use mostly telephone interviewing instead of the more costly face-to-face interviews of previous surveys in the FHWAR series. That requires shortening of the questionnaires, elimination of the map booklets and flashcards, and other changes in the FHWAR methodology. In addition, the 1991 survey is designed to provide a limited test of the difference between estimates based on a four-month recall period and a one-year recall period as a further measure of recall error.

Additional empirical research might focus on differences in outcomes between episodic recall and estimation strategies. This research would address the conditions under which respondents provide overestimates and those under which they provide underestimates. More specifically, the research would examine the extent to which the direction of the error is related to the use of episodic recall or estimation, and the extent to which the direction of error is related to social desirability or prestige bias.

7. References

Cannell, C. (1965). Reporting of Hospitalization in the Health Interview Study. Vital and Health Statistics, NCHS, 2(6), Public Health Service.

Cash, W.S. and Moss, A.J. (1969). Methodology Study for Determining the Optimum Recall Period for the Reporting of Motor Vehicle Accidental Injuries. Proceedings of the Social Statistics Section, American Statistical Association, 364–369.

Chase, D.R. and Godbey, G.C. (1983). Research Note: The Accuracy of Self-Reported Participation Rates. Leisure Studies, 2, 231–235.

Chase, D.R. and Harada, M. (1984). Response Error in Self-Reported Recreation Participation. Journal of Leisure Research, 16, 322–329.

Cohen, B., Erickson, P., and Powell, A. (1984). The Impact of Length of Recall Period on the Estimation of Health Events. Proceedings of the Social Statistics Section, American Statistical Association, 593–598.

Ferber, M.A. and Birnbaum, B.G. (1979). Retrospective Earnings Data: Some Solutions for Old Problems. Public Opinion Quarterly, 43, 112–118.

Gems, B., Ghosh, D., and Hitlin, R. (1982). A Recall Experiment. Impact of Time on Recall of Recreational Fishing Trips. Proceedings of the Section on Survey Research Methods, American Statistical Association, 372–375.

Ghosh, D.N. (1978). Optional Recall for Discrete Events. Proceedings of the Section on Survey Research Methods, American Statistical Association, 615–616.

Gray, P.G. (1955). The Memory Factor in Social Surveys. Journal of the American Statistical Association, 50, 344–363.

Jaeger, C. and Pennock, J. (1962). A Note on Accuracy and Consistency of Response in Household Surveys. Unpublished memorandum, Agricultural Research Service, U.S. Department of Agriculture.

Means, B., Mingay, D.J., Nigan, A., and Zarrow, M. (1988). A Cognitive Approach to Enhancing Health Survey Reports of Medical Visits. In M.M. Gruneberg, P.E. Morris, and R.N. Sykes (eds.), Practical Aspects of Memory: Current Research

and Issues. Chichester, John Wiley.

Metz, J.F. (1956). Accuracy of Response Obtained in a Milk Consumption Study. Methods of Research in Marketing, Paper No. 5. Ithaca, N.Y.: Cornell University College of Agriculture.

Murphy, L.R. and Cowan, C.D. (1976). Effects of Bounding on Telescoping in the National Crime Survey. Proceedings of the Social Statistics Section, American Statistical Association, 633–638.

National Center for Health Statistics (1962). Volume of X-Ray Visits. Health Statistics (PHS Pub. no. 584-B38), Washington, D.C.: Public Health Service.

National Center for Health Statistics (1965a). Comparison of Hospitalization Reporting. Vital and Health Statistics (PHS Pub. no. 1000, series 2, no. 8), Washington, D.C.: Public Health Service.

National Center for Health Statistics (1965b). Health Interview Responses Compared with Medical Records. Vital and Health Statistics (PHS Pub. no. 1000, series 2, no. 7), Washington, D.C.: Public Health Service.

National Center for Health Statistics (1967). Interview Data on Chronic Conditions Compared with Information Derived from Medical Records. Vital and Health Statistics (PHS Pub. no. 1000, series 2, no. 23), Washington, D.C.: Public Health Service.

Neter, J. and Waksberg, J. (1964). A Study of Response Errors in Expenditures Data from Household Interviews. Journal of the American Statistical Association, 59, 18–55.

Politz, A. (1958). Description of Operational Design and Procedures. Life Study of Consumer Expenditures, 4, Time, Inc.

Sudman, S. (1964). On the Accuracy of Recording of Consumer Panels: I. Journal of Marketing Research, 1, 14–20.

Sudman, S. and Ferber, R. (1970). Experiments in Obtaining Consumer Expenditures by Diary Methods. University of Illinois, Report to BLS, U.S. Department of Labor.

Tracy, P.E. and Fox, J.A. (1981). The Validity of Randomized Response for Sensitive Measurements. American Sociological Review, 46, 187–200.

Waksberg, J., and Valliant, R. (1978). Final Report on the Evaluation and Calibration of NEISS. Westat, Inc. for Consumer Products Safety Commission.

Westat, Inc. (1989). Investigation of Possible Recall/Reference Period Bias in National Surveys of Fishing, Hunting and Wildlife-Associated Recreation. Final Report submitted to U.S. Department of Interior, Fish and Wildlife Service.

Woltman, H., Bushery, J., and Carstensen, L. (1975). Recall Bias and Telescoping in the National Crime Survey (NCS). Internal Memorandum, U.S. Bureau of the Census.

Woolsey, T.D. (1953). Results of the Sick-Leave Memory Test of October, 1952. U.S. Department of Health, Education, and Welfare, unpublished memorandum.

Wyner, G.A. (1980). Response Errors in Self-Reported Number of Arrests. Sociological Methods and Research, 9, 161–177.

Received April 1991
Revised January 1992