RECALL STRATEGIES FOR ESTIMATION OF SMOKING LEVELS IN HEALTH SURVEYS[1]

Barbara Means & Gary E. Swan, SRI International
Jared B. Jobe & James L. Esposito, National Center for Health Statistics
Elizabeth F. Loftus, University of Washington

INTRODUCTION

Many survey questions ask about the frequency of specific behaviors (e.g., consumption of a particular food, use of a product, voting). The implicit assumption has been that responses to these items are formed through a process of episodic recall, in which the respondent mentally retrieves all pertinent incidents and then tallies them. Given this assumption, survey methodologists (e.g., Bradburn, 1983) have focused their attention on potential sources of response error associated with episodic recall (i.e., retrieval failures and telescoping effects).

Recently, this assumption has been challenged, particularly in cases where the target behavior is high in frequency (Blair & Burton, 1986). The smoking frequency item appearing on most smoking surveys, "How many cigarettes do you smoke a day?" clearly places heavy demands on the smoker trying to episodically recall each cigarette smoked during a given day. Given not only the frequency with which smokers light a cigarette, but also the almost unconscious nature of this activity, it seems reasonable to assume that few subjects answer smoking frequency questions through exhaustive episodic recall of individual incidents. If not episodic recall, then what processes are respondents using and what is the relative accuracy of different frequency estimation strategies?

Blair and Burton (1986) explored the issue of alternative frequency estimation strategies in a recent series of studies. They administered survey questions about frequency of behaviors such as dining in a restaurant or using an automatic teller machine. Immediately after answering the frequency question, the respondent was queried about the method used to arrive at the response. Episodic recall was the most frequently reported strategy when there were three or fewer incidents in the event category, but use of this strategy declined rapidly as event frequency rose. For those who ate in a restaurant 10 times or more, for example, no one reported basing their frequency estimate on episodic recall. Instead, responses were based on what Blair and Burton called "direct" estimates or on procedures for decomposing the reference period and computing rates. For a very frequent, commonplace behavior, writing checks, nearly two-thirds of their subjects reported using direct estimates under normal interviewing procedures. The psychological literature suggests that such direct estimates are subject to biases quite different from the omission and telescoping errors found in episodic recall. Tversky and Kahneman (1982) suggest that a commonly used heuristic for producing these estimates is availability judgments. Easy-to-retrieve events are assumed to have a high rate of occurrence. Because factors other than frequency affect retrievability, this heuristic can lead to bias. For example, in one Tversky and Kahneman study, subjects were asked to consider a letter (K, L, N, R, or V) and asked whether the letter is more likely to appear in the first position or in the third position in English words. Although these letters are in fact more likely to appear in the third position, it is easier to retrieve words starting with a given letter than words in which the letter appears in the third position. Each of the letters was judged by the majority of subjects to be more frequent in the first position of words than in the third position and the bias was quite strong: the mean estimate for the ratio of the two frequencies was 2:1. Given the relationship between event salience and retrievability and the low salience of individual smoking episodes, the Tversky and Kahneman work would suggest that using an availability strategy would tend to lead to under-estimates of smoking rate.

On the other hand, a study reported by Bruce and Read (1988) would suggest that however biased direct estimates based on availability may be, they are better than estimates based on attempted episodic recall for high-frequency events. In that study (conducted with F. Craik), subjects saw lists of words from different semantic categories (e.g., animals). After viewing a list, subjects were asked first to give an initial frequency estimate (presumably based on availability) for each category (e.g., "How many animals were on the list?"). This first frequency estimate was followed by cued recall: the experimenter provided category labels and instructed the subject to recall as many instances of that category from the word list as he or she could. This cued recall was followed by a second frequency estimate. These second estimates were typically less accurate than the initial estimates, the interpretation being that although subjects had failed to recall all instances, they used their partial recall as a guide in making the second estimate.

Another approach to estimating frequency reported by subjects interviewed by Blair and Burton (1987) and in a similar study reported by Bradburn, Rips, and Shevell (1987) entails decomposing the reference period, computing a frequency for one or more parts of it, and then multiplying or adding part frequencies to arrive at an estimate for the whole. Although subjects have reported using this kind of strategy in several studies, evidence concerning the relative accuracy of frequency estimates derived in this way is lacking.

The study reported here is a first step toward (1) identifying the strategies that people do use to answer questions about smoking frequency and (2) assessing the relative accuracy of different frequency estimation strategies. To permit assessment of recall accuracy, subjects collected behavioral evidence of their smoking behavior for 4-5 days. Afterwards, they were given an unexpected question about the number of cigarettes they had smoked on one of the days during that assessment period. Using a four-group, between-subjects design, we allowed one group of subjects to answer the frequency question using any method they liked while constraining the other three groups to employ a particular frequency estimation strategy.

METHODS

Local advertisements solicited smokers to participate in the research for a $50 fee. Potential subjects were told that the study concerned different methods for measuring nicotine dosage, including

biochemical analyses of saliva samples. Individuals responding to the advertisement were asked an open-ended question about how much they smoke per day and were screened to make sure that they were not actively trying to quit smoking at that time. A sample of 25 subjects, 10 males and 15 females, was obtained in this manner.

Subjects participated in an initial face-to-face interview, collected behavioral evidence of their weekday smoking for 4-5 days, and then returned for a second interview during which they were asked about smoking frequency on a target day and tested for saliva cotinine (a chemical byproduct of nicotine).

The purposes of the initial interview were to (1) give subjects instructions for collecting the behavioral evidence and (2) focus their attention on aspects of their smoking behavior other than frequency. The latter purpose was accomplished by inspecting the subject's cigarette brand, asking questions about smoking style features (such as depth of inhalation), and having the subject smoke a cigarette in the laboratory under the researcher's observation.

Subjects were asked to sign a form pledging to collect all of their butts in the provided containers for each day, putting all Monday butts in the container labeled "Monday," etc. They were instructed also to keep a record as each cigarette was lit.

After collecting this behavioral evidence of their smoking for a period of 4 or 5 days (depending upon the interview schedule), the subjects brought their behavioral evidence to the laboratory and participated in a second interview. The critical question in this interview was a query concerning the number of cigarettes they had smoked on one of the days during the assessment period (the target day). Depending upon their treatment group, subjects were either trained in a specific frequency estimation strategy and directed to use it in answering the question or allowed to compute a smoking frequency in any way they liked. The three specified strategies, based upon episodic recall, availability, and decomposition, are described in Table 1.

In the free-strategy/think-aloud condition, subjects were not told how they should answer the smoking frequency question, but were directed to think aloud as they formulated their answer. All subjects gave a confidence estimate for the smoking frequency they reported and then gave a saliva sample for biochemical analysis.

Table 1. Frequency Estimation Strategies

Additive Decomposition

The typical weekday is divided into different types of activities such as "commuting," "in the office," and "after dinner." The subject is asked to estimate the number of cigarettes consumed in each of these portions of the day. After all separate estimates are made, they are summed to arrive at a total day's estimate.

Availability

The subject is instructed to answer the smoking frequency question quickly with a "gut reaction" without trying to think of specific instances of smoking.

Episodic Recall

The subject is instructed to take a selected weekday and to work through the day recalling all instances of smoking. The subject is then asked to use the recalled instances to arrive at a daily total.

RESULTS

Self Reports Prior to Participation

Subject responses to the telephone screening question about smoking level were reviewed to get a sense of the precision of their prior knowledge about how much they smoke. Slightly more than half of the subjects (13 or 52%) described their smoking in terms of packs rather than number of cigarettes. Of those subjects who did respond in terms of cigarettes, 3 gave a range; 2 spoke of "more than 20" or "less than 20"; and the remainder gave a number that was a multiple of 10 (most, in fact, were multiples of 20). Hence, we concluded that subjects have a rough idea of their own smoking level but do not demonstrate precise knowledge. The digit bias observed in other studies of smoking (Pechacek, Fox, Murray, & Luepker, 1984) is apparent here as well.

Quality of Behavioral Evidence

The two forms of behavioral evidence -- number of cigarette butts saved and cigarette lighting record -- were compared for each day within the assessment period. A high level of agreement ($r = .91$) was found. For the target day, the number of cigarettes on the behavioral record was identical to the number of cigarette butts in the container for 15 of the 25 subjects. Where differences did occur, they were small in magnitude, averaging just 0.64 cigarette. For simplicity's sake, we report here only those analyses using the behavioral record as the "true" smoking frequency.

Smoking Levels

The mean number of cigarettes smoked during the days when subjects kept records was 22.94 and the median was 20.20. The lightest smoker averaged 7.60 cigarettes per day, and the heaviest averaged 58.

Looking at the mean daily smoking by treatment condition, the mean was 28.57 for the episodic strategy group, 20.17 for additive decomposition, 22.67 for availability, and 23.17 for the free-strategy/think-aloud group. Thus, the means for target-day smoking were quite close with the exception of the somewhat higher smoking level among episodic subjects.

Accuracy of Frequency Reports

Table 2 shows the mean difference between the number of cigarettes reported and the number on the behavioral record for the target day in each of the treatment conditions. We tested whether each of these means was significantly different from zero. The only significant difference was found in the availability condition, $t(6) = -4.58$, $p < .01$, in which reports averaged 3.50 cigarettes lower than the behavioral record. In the free-strategy/think-aloud condition, reports averaged an insignificant 1.17 cigarettes below the behavioral record. In contrast to these two conditions, subjects in the episodic and additive decomposition conditions tended to overreport their smoking for the target day, but the differences were not significant.

The third column in Table 2 shows the mean absolute difference between the number of cigarettes reported and the number in the behavioral record ("gross

error"). Gross error was largest for the episodic and additive decomposition conditions (4.43 and 4.17 cigarettes, respectively) and smallest for the free-strategy/think aloud condition (1.50). Gross error rates were significant (different from 0) in all conditions except additive decomposition. When tested through a one-way ANOVA, however, the effect of condition on gross error rate was not significant.

A third measure of accuracy was created in order to correct for the fact that individuals with different levels of smoking had recall tasks of different difficulty. This measure, (Behavioral Record Count-Reported Frequency) / Behavioral Record Count is shown in the fourth column of Table 2. An ANOVA testing for the effect of treatment condition on this measure was significant, $F(3,21) = 3.70$, $p<.05$. Comparisons between conditions revealed that performance in the availability condition differed significantly from each of the other conditions. Thus, reported frequencies in the availability condition are a smaller proportion of the corresponding record frequencies than in any of the other strategy conditions.

Free-Strategy/Think-Aloud Strategies

The strategies that subjects reported using in the free-strategy/think-aloud condition were examined. Three subjects said that they used "knowledge about my usual daily smoking", a strategy based upon semantic rather than episodic memory. Such knowledge may be tested against memory for episodes on the target day: "My average was about 20 and I cannot recall Thursday as being a day I smoked less or more." Three other subjects reported complex strategies that involved dividing the day up into segments (decomposition) and then either recalling all the individual events during each segment (episodic) or using the first number that came to mind for each segment (availability). The protocol for one of the subjects using this kind of strategy illustrates the approach: "I think it was about 11. I have 3 in the car coming to work and 3 going home, 1 at each break, 2 at lunchtime, 1 at home." A piece of converging evidence for these subject self reports was obtained by having a researcher independently review the verbal protocols for a sample of 19 subjects. Of these, 15 protocols were sufficient to permit categorizing the subject's strategy into one of six classifications. Among these, 11 of the experimenter classifications (73%) matched the subject's self-reported strategy.

Cotinine

Saliva samples were analyzed for cotinine, a nicotine byproduct. Cotinine levels in all subjects were in the range normally associated with smokers. Correlations between cotinine level and the various behavioral evidence of smoking ranged between .35 and .40. Thus, although cotinine is the preferred biochemical measure for distinguishing between smokers and nonsmokers (Pechacek et al., 1984) it does not appear to be calibrated accurately enough to replace self-report as the primary measure of smoking frequency among smokers.

Confidence

Subjects rated their confidence in their frequency estimates on a five-point scale, with 5 = "exactly correct" and 1 = "off by more than 10 cigarettes". Table 3 shows the confidence ratings by treatment condition. The means suggest that subjects are least confident about frequency estimates derived through attempts to recall all smoking events during the target day (episodic strategy) and most confident about frequencies derived through additive decomposition. It is interesting to note that the pattern of differences in the confidence ratings is not congruent to that for recall accuracy (where the trend was for better recall in the episodic condition than with additive decomposition).

Table 3. Mean Subject Confidence

Condition	Mean Rating
Availability	3.33
Additive Decomposition	3.67
Episodic	2.29
Free	3.50

Key to Confidence Ratings:
1 = Off by more than 10 cigarettes
2 = Off by no more than 10 cigarettes
3 = Off by no more than 5 cigarettes
4 = Off by no more than 1 or 2 cigarettes
5 = Exactly correct

DISCUSSION

Of the treatment conditions in this study, the availability condition, in which subjects are urged to respond quickly with a gut reaction, may be most likely to elicit the processes typically used in answering survey questions about smoking frequency. This is because the availability condition comes closest to the rapid pace used in standard national surveys. Although standard surveys resemble the free-strategy/think-aloud condition in not instructing subjects to use any particular method for estimating frequency, we would argue that their rapid pace discourages subjects from using the more deliberate strategies observed in this study when subjects are urged to think aloud (Lessler, Salter, & Tourangeau, 1989). Subjects in the free-strategy/think-aloud condition averaged 5.25 seconds to respond to the frequency question while those in the availability condition had a mean time of 3.25 seconds. If this assumption about the similarity of the availability condition to standard survey procedures is correct, the fact that the availability condition led to significant underreporting of smoking frequency has important implications for national surveys on smoking and other high-frequency behaviors.

Performance in the free-strategy/think-aloud and the episodic conditions was considerably better than that in the availability condition; neither group had reported frequencies significantly different from their behavioral records. The fourth condition, additive decomposition, tended to produce overreporting of smoking frequencies. The failure to find a significant difference between reported frequency and behavioral record in this strategy condition appears to be the result of a large variance coupled with a small sample size. Results for this condition are important in light of recently reported findings that subjects use additive decomposition strategies when answering survey questions about high-frequency events. The large variance observed in this condition coupled with the fact that two different kinds of additive decomposition strategies were used in the free-strategy/think-aloud condition, suggests that subjects in the additive decomposition

group were not homogeneous in their processing. The large variance observed in our condition might be attributable at least in part to some subjects coupling the decomposition strategy with episodic recall of events within each part while other subjects employ an availability heuristic.

A follow-up study with a much larger sample size is now being conducted to replicate and extend the findings summarized in this paper. In particular, we are using a sample large enough to permit exploring the issue of whether the strategies people choose when unconstrained and the most effective strategies vary depending upon the frequency level of the behavior. Episodic strategies, for example, may be quite useful for people who are light smokers but ineffective for heavy smokers. The larger study will explore the frequency by strategy interactions and the biases associated with each combination of strategy and smoking level.

REFERENCES

Blair, E., & Burton, S. (1986). Processes used in the formulation of behavioral frequency reports in surveys. American Statistical Association, Proceedings of the Section on Survey Methods Research, 481-487.

Bradburn, N. M. (1983). Response effects. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), Handbook of survey research (pp. 289-328). New York: Academic Press.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. Science, 198, 157-161.

Bruce, D., & Read, J. D. (1988). The how and why of memory for frequency. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory (pp. 317-322). Chichester, UK: Wiley.

Lessler, J., Salter, W., & Tourangeau, R. (1989). Questionnaire design in the cognitive research laboratory. Vital and Health Statistics, Series 6, No. 1 (DHHS Publication No. PHS 89-1076). Washington, DC: U.S. Government Printing Office.

Pechacek, T. F., Fox, B. H., Murray, D. M., & Luepker, R. V. (1984). Review of techniques for measurement of smoking behavior. In J. D. Matarazzo, S. M. Weiss, J. A. Herd, & N. E. Miller (Eds.), Behavioral health: A handbook of health enhancement and disease prevention (pp. 729-754). New York: Wiley.

Tversky, A., & Kahneman, D. (1982). Availability: A heuristic for judging frequency and probability. In D. Kahneman, & P. Slovic (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 163-178). Cambridge: Cambridge University Press.

Footnote

[1] This research is being conducted under DHHS contract 200-88-7074 between NCHS and SRI International, and is part of a larger project supported by grant SES 86-12320 from the National Science Foundation to the National Center for Health Statistics.

Table 2. Mean Level of Smoking in Behavioral Records and Self Reports

Condition	Target Day Record	Record-Reported	Absolute Value Record-Reported	Record-Reported / Record
Availability	22.67 (10.46)	3.50* (1.87)	3.50* (1.87)	0.15* (0.05)
Additive Decomposition	20.17 (6.59)	-3.50 (5.89)	4.17 (5.34)	-0.16 (0.27)
Episodic	28.57 (17.44)	-1.57 (5.59)	4.43* (3.36)	-0.12 (0.23)
Free Strategy/ Think Aloud	23.17 (8.57)	1.17 (1.60)	1.50* (1.22)	0.05 (0.09)

Note: A minus sign indicates overreporting. Standard deviations appear in parentheses.

* $p < .05$.