Epidemiology

Volume 5, Number 1

January 1994

Published jointly by
Williams & Wilkins
and Epidemiology Resources Inc.
Nancy A. Dreyer, President

Editorial Office Address:
Kenneth J. Rothman, Editor
Epidemiology
One Newton Executive Park
Newton Lower Falls, MA 02162

Business Office Address:
Epidemiology
Williams & Wilkins
428 East Preston Street
Baltimore, MD 21202

EDITORIALS

Improving Long-Term Recall in Epidemiologic Studies

Epidemiologic research relies heavily on retrospective reporting of exposures to putative risk factors. Many diseases are believed to have long latency periods, and, consequently, exposures in the distant past are often considered the most important for understanding disease etiology. Retrospective studies are believed to be methodologically weaker than prospective studies, mainly because exposure data must be obtained after the exposures have occurred, and the memory of previous exposures may be influenced by random and systematic errors and bias. Retrospective studies, however, still afford several advantages over prospective studies. The potential for improving the quality of recalled exposure data deserves the close attention of epidemiologists.

Despite concerns regarding the validity and reproducibility of retrospective reporting, relatively few epidemiologic investigations have specifically examined the accuracy[1-6] or reproducibility of long-term (≥10 years) recall.[7-15] The studies conducted to date have examined the long-term recall of anthropometric and maturation information,[3-6] smoking and alcohol consumption history,[2,15] dietary intake,[1,7-13] and physical activity.[14,15] In general, recall accuracy and reproducibility were best for the reporting of past anthropometric measurements, smoking, and alcohol consumption, but were weaker for less salient items such as dietary intake and physical activity. The level of agreement between originally reported and recalled data has ranged from negative correlations for certain dietary items reported 44 years later[13] to high positive correlations for body size measurements.[3-5] Lee *et al*[15] have shown that intersubject variance in recall is largest for smoking exposure data, intermediate for alcohol, and lowest for physical activity, but that recall error is greatest for physical activity and lowest for smoking. These epidemiologic studies[1-15] have used standard interviewing methods for the reporting of past exposures; no attempt was made to enhance the recall ability of the study subjects.

The paper by Cumming and Klineberg[16] in this issue adds to this literature. These investigators found acceptable reproducibil-

ity of long-term recall up to 82 years ago for weight, height, reproductive factors, smoking, and a few dietary intakes examined. Past physical activity was poorly recalled, and recall reproducibility decreased with cognitive impairment, increasing age, and male gender.

A few epidemiologic studies have examined variables that include age, sex, body mass index, education, income, and life-style choices as determinants of long-term recall ability.[2,3,5,13,14,17] Drawing conclusions from these studies regarding the predictors of long-term recall is hampered by the fact that few studies have been conducted, and each has used different methods for examining recall accuracy and reproducibility. Overall, none of these demographic and personal factors has been consistently shown to be a predictor of recall ability. Some studies have found poorer recall associated with male gender,[3,13] advanced age,[17] low education, and income,[17] whereas others have found that sex, age, or other demographic and personal characteristics do not influence recall ability.[2,14] One study found a systematic bias in recalled weight according to body mass index: lean subjects overestimated and heavy subjects underestimated their weight over all time periods.[5] Other variables likely to influence recall ability have not been examined in epidemiologic studies. Obvious factors that should be considered are the salience, nature, and sensitivity of the information being recalled, the data collection methods themselves (for example, questionnaire wording, order, nature and level of detail of the questions, type of administration, context for interview, memory probes), and the length of the recall time interval.

Understanding how information is recalled, what factors predict accurate and reliable recall, and how recall can be improved are the areas that now require the attention of epidemiologists conducting retrospective research. Consideration of these methodologic issues will mitigate the influence of measurement errors and biases possible with retrospective data.

During the past 15 years, investigators in the fields of survey research and cognitive psychology have begun interdisciplinary work to improve survey methodology.[18] Researchers in these two fields overcame the differences in their methodologies to undertake mutually beneficial collaborative research. Although this research has been directly applied to health surveys, retrospective epidemiologic investigations, particularly case-control studies, have yet to adopt the cognitive survey methods that have been proposed and successfully used by this body of researchers.

Cognitive psychologists provided a theoretical framework for how survey questions are answered. Using this conceptualization of how cognitive processes work has allowed survey researchers to design and test improved measurement instruments. The basic theoretical model involves four stages of cognitive processes undertaken by respondents when answering a question. These stages are question comprehension, information retrieval, estimation and judgment, and response formulation.[19] During the first stage, the respondents interpret the meaning of the question. In the second stage, they search long-term memory for the relevant information. The third stage occurs when respondents evaluate the information retrieved from memory and decide whether it is relevant and adequate. At this stage, if the respondents decide the information is adequate, a response is formulated, or, alternatively, if the information is deemed inadequate, they may initiate another search of their memory. In the final stage, the respondents decide what answer to provide by weighing several factors. Although described here as a sequential process, the entire sequence has been hypothesized to be quite flexible, with numerous control processes (decision and judgment) occurring before and after retrieval of information from memory.[20]

At each of these four stages of the cognitive process, the possibility exists that personal and methodologic factors will introduce reporting errors and bias. During question comprehension, factors such as age, sex, education, intelligence, ethnicity, disease status, and personal experiences could influence the ability of the respondent to answer the question.[21] When retrieving the information from memory and evaluating whether it is correct, the determinants of recall ability are likely to be the time interval since the event, the type of information being recalled (episodic or generic information), the amount of detail in the question, the salience of the subject matter to the respondent, the length of the reference period, and the frequency and regularity of the target experience or exposure.[22] Finally, during response formulation, factors that are weighed by the respondent will include the sensitivity of the question, the social desirability of a particular response, the perceived "correct" response, and the probable accuracy of the answer.[23]

Cognitive psychologists and survey researchers are studying each of these four stages of cognitive processes and are devising and evaluating methods to improve recall and reduce reporting errors and bias. These researchers have found that question comprehension can be improved by increasing the question length, by providing more instructions, by using simpler wording, by changing the question order to be

more compatible with autobiographical memory (asking first *why*, then *how*, and, finally, *when*), and by using an interview-administered questionnaire.[21-23] A vital step in improving question comprehension has been to test the questionnaire first in a laboratory setting and then in a field pre-test.[24]

To improve information retrieval, cognitive interviewing techniques have been applied.[23] These techniques include such methods as *concurrent think-aloud interviews* (respondents think aloud when answering questions, and responses are extensively probed), *retrospective think-aloud interviews* (respondents answer all questions first, then are asked how they arrived at their answers), *paraphrasing* (respondents repeat the questions in their own words), *probes* (follow-up questions used to gain more information about respondents' strategies for answering questions), and *memory cues* (interviewer reads terms which are intended as aids to recall). Recall has also been shown to be improved by constructing a *personal time line* that permits subjects to date events/exposures more accurately relative to landmark events that they can remember with high degrees of accuracy (for example, weddings, birthdays, vacations). Recall is also improved if subjects are encouraged to provide details that may help them remember an event.[21] All of these cognitive interviewing techniques assist respondents in remembering by generating a context for the events/exposures in their past that they are being asked to recall.

The highly standardized interviewing methods that have been used repeatedly in epidemiologic studies have been shown to be less effective than cognitive interviewing techniques in recalling accurate and reliable information.[21] An experimental method that used cognitive interviewing techniques was shown to be more effective than a conventional interview in helping subjects recall an accurate smoking history for a distant time period.[21] Experimental interview methods are particularly useful for medium- or low-salience events/exposures, since respondents experience greater difficulty in recalling these past data. Standardized interviews reduce the ability and opportunity of subjects to think aloud, ask additional questions, respond to memory cues or probes, construct a personal time line, etc. In short, these interviews do not permit the type of discussion between interviewer and respondent shown to enhance recall ability.[21]

When retrieving and estimating information, several sources and types of error may affect the final answer.[22,25] These include *telescoping* (failure to recall exactly when an event occurred), *reconstruction* (merging two similar events into one), substituting *generic* memory for *episodic* memory (providing usual patterns rather than actual specifics of an event/exposure), and making errors in *dates*. Problems with telescoping can be reduced by using bounded interviews and landmark events, that is, time lines. Errors associated with dates can be minimized by asking about the dates after obtaining all other information on an event/exposure. To overcome problems with reconstruction, information on specific episodes can be retrieved by using *decomposition* (providing more retrieval cues and increasing the respondent's time on the task), time lines, and a personal calendar. For some exposures, such as usual diet, physical activity, or other life-style activities, generic memory may actually be more important and relevant than episodic memory.[26]

In the final cognitive stage in recall and reporting, errors and biases can be reduced, and the accuracy and reliability of the information reported can be improved by motivating the subjects to provide honest answers and by providing them with positive feedback and sufficient time to formulate their responses.[22] The interviewer should remain neutral and nonjudgmental throughout the interview to decrease the possibility of biased responses. *Confidence ratings* can also be used at this stage to assess the degree of confidence that the respondents have in the accuracy of their responses.[23] The *response latency*, that is, the length of time elapsed between asking the question and receiving a response, can also be measured.[23] These last two measurements provide additional information on the level of accuracy and reproducibility of the response.

Epidemiologists have yet to incorporate systematically methods for improving recall accuracy and reproducibility into the data collection procedures used in retrospective studies. Research should be focused on studying the predictors of recall ability to devise methods that reduce recall errors and bias. If these predictors are demographic characteristics associated with particular subgroups of the population with specific recall problems, then different interviewing techniques customized to each subgroup's needs could be developed. A large component of recall errors and bias will likely be attributable to aspects of the questionnaire design and interviewing methods. Adopting the cognitive interviewing techniques, questionnaire design, and field pre-testing strategies, as developed by cognitive psychologists and survey researchers, should result in improved data quality in epidemiologic studies. All of these methods will require more preparation, testing, and interviewing time. The clear benefits, however, are the improved validity and reproducibility of retrospective data, which should be a principal objec-

tive of any researcher relying on data recalled from the distant past.

Christine M. Friedenreich

Department of Community Health Sciences,
The University of Calgary,
3330 Hospital Drive N.W., Calgary, Alberta,
Canada T2N 4N1 (address for correspondence)

References

1. Sobell J, Block G, Koslowe P, Tobin J, Andres R. Validation of a retrospective questionnaire assessing diet 10–15 years ago. Am J Epidemiol 1989;130:173–187.
2. Krall EA, Valadian I, Dwyer JT, Gardner J. Accuracy of recalled smoking data. Am J Public Health 1989;79:200–206.
3. Casey VA, Dwyer JT, Coleman KA, Krall EA, Gardner J, Valadian I. Accuracy of recall by middle-aged participants in a longitudinal study of their body size and indices of maturation earlier in life. Ann Hum Biol 1991;18:155–166.
4. Must A, Willett WC, Dietz WH. Remote recall of childhood height, weight, and body build by elderly subjects. Am J Epidemiol 1993;138:56–64.
5. Stevens J, Keil JE, Waid LR, Gazes PC. Accuracy of current, 4-year, and 28-year self-reported body weight in an elderly population. Am J Epidemiol 1990;132:1156–1163.
6. Damon A, Bajema CJ. Age at menarche: accuracy of recall after thirty-nine years. Hum Biol 1974;46:381–384.
7. Wu ML, Whittemore AS, Jung DL. Errors in reported dietary intakes. II. Long-term recall. Am J Epidemiol 1988;128:1137–1145.
8. Bakkum A, Bloemberg B, van Staveren WA, Verschuren M, West CE. The relative validity of a retrospective estimate of food consumption based on a current dietary history and a food frequency list. Nutr Cancer 1988;11:41–53.
9. Thompson FE, Lamphiear DE, Metzner HL, Hawthorne VM, Oh MS. Reproducibility of reports of frequency of food use in the Tecumseh Diet Methodology Study. Am J Epidemiol 1987;125:658–671.
10. Jensen OM, Wahrendorf J, Rosenqvist A, Geser A. The reliability of questionnaire-derived historical dietary information and temporal stability of food habits in individuals. Am J Epidemiol 1984;120:281–290.
11. Byers TE, Rosenthal RI, Marshall JR, Rzepka TF, Cummings KM, Graham S. Dietary history from the distant past: a methodological study. Nutr Cancer 1983;5:69–77.
12. Lindsted KD, Kuzma JW. Long-term (24-year) recall reliability in cancer cases and controls using a 21-item food frequency questionnaire. Nutr Cancer 1989;12:135–149.
13. Dwyer JT, Gardner J, Halvorsen K, Krall EA, Cohen A, Valadian I. Memory of food intake in the distant past. Am J Epidemiol 1989;130:1033–1046.
14. Blair SN, Dowda M, Pate RR, Kronenfeld J, Howe HG, Parker G, Blair A, Fridinger F. Reliability of long-term recall of participation in physical activity by middle-aged men and women. Am J Epidemiol 1991;133:266–275.
15. Lee MM, Whittemore AS, Jung DJ. Reliability of recalled physical activity, cigarette smoking, and alcohol consumption. Ann Epidemiol 1992;2:705–714.
16. Cumming RG, Klineberg RJ. A study of the reproducibility of long-term recall in the elderly. Epidemiology 1994;5:116–119.
17. Kuzma JW, Lindsted KD. Determinants of long-term (24-year) diet recall ability using a 21-item food frequency questionnaire. Nutr Cancer 1989;12:151–160.
18. Jobe JB, Mingay DJ. Cognition and survey measurement: history and overview. Appl Cognit Psychol 1991;5:175–192.
19. Tourangeau R. Cognitive science and survey methods. In: Jabine TB, Straf ML, Tanur JM, Tourangeau R, eds. Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines. Washington DC: National Academy Press, 1984;73–100.
20. Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. Appl Cognit Psychol 1991;5:251–267.
21. Means B, Swan GE, Jobe JB, Esposito JL. An alternative approach to obtaining personal history data. In: Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S, eds. Measurement Errors in Surveys. New York: John Wiley and Sons, 1991;167–183.
22. Jobe JB, Tourangeau R, Smith AF. Contributions of survey research to the understanding of memory. Appl Cognit Psychol (in press).
23. Jobe JB, Mingay DJ. Cognitive research improves questionnaires. Am J Public Health 1989;79:1053–1055.
24. Bercini DH. Pretesting questionnaires in the laboratory: an alternative approach. J Exp Anal Envir Epidemiol 1992;2:241–248.
25. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. Science 1987;236:157–161.
26. Smith AF, Jobe JB, Mingay DJ. Retrieval from memory of dietary information. Appl Cognit Psychol 1991;5:269–296.

Physical Exertion and Reproductive Success

Two articles in this issue of **Epidemiology** address the topic of physical exertion and its effects on reproductive success in women.[1,2] With the exception of certain categories of exertion, both studies largely found no evidence that physical exertion deleteriously affected ability to conceive or continue a pregnancy. Both studies also illustrate some of the difficulties of performing epidemiologic research on this topic.

The article by Florack *et al*[1] addresses the influence of occupational activity on cycle characteristics and fecundability. This prospective study assessed ability to conceive in women employed in a number of Dutch hospitals who indicated their intent of becoming pregnant soon. Menstrual detail was assessed retrospectively. Physical exertion was rated on a "fatigue score" based upon total energy expenditure during the day and an "intensity score" based upon peak levels of activity. Although fatigue and intensity scores did not