Vital and Health Statistics

Autobiographical Memory for Health-Related Events

Series 6: Cognition and Survey Measurement No. 2

This report explores the hypothesis that recurring events lead to creating generic memories and to increased difficulty in recalling individual incidents. It describes research using cognitive theory as a basis for designing a two-part intervention procedure for facilitating recall of personal events.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Center for Health Statistics

Hyattsville, Maryland
August 1989
DHHS Publication No. (PHS) 89-1077

Copyright information

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Suggested citation

Means B, Nigam A, Zarrow M, et al. Autobiographical memory for health-related events. National Center for Health Statistics. Vital Health Stat 6(2).1989.

Library of Congress Cataloging-in-Publication Data

Autobiographical memory for health-related events.
(Vital & health statistics. Series 6, Cognition and survey measurement ; no. 2) (DHHS pub. ; no. (PHS) 89-1077)
Bibliography: p.
1. Health surveys. 2. Memory. 3. Cognition.
I. Means, Barbara II. National Center for Health Statistics (U.S.) III. Series.
IV. Series: DHHS publication ; no. (PHS) 89–1077. [DNLM: 1. Awareness.
2. Health Surveys. 3. Memory. 4. Research Design. WL 102 A939]
RA408.5.A94 1989 614.4'2 89–600018
ISBN 0–8406–0413–0

National Center for Health Statistics

Manning Feinleib, M.D., Dr.P.H., *Director*

Robert A. Israel, *Deputy Director*

Jacob J. Feldman, Ph.D., *Associate Director for Analysis and Epidemiology*

Gail F. Fisher, Ph.D., *Associate Director for Planning and Extramural Programs*

Peter L. Hurley, *Associate Director for Vital and Health Statistics Systems*

Stephen E. Nieberding, *Associate Director for Management*

Charles J. Rothwell, *Associate Director for Data Processing and Services*

Monroe G. Sirken, Ph.D., *Associate Director for Research and Methodology*

Office of Research and Methodology

Monroe G. Sirken, Ph.D., *Associate Director*

Kenneth W. Harris, *Special Assistant for Program Coordination and Statistical Standards*

Lester R. Curtin, Ph.D., *Chief, Statistical Methods Staff*

James T. Massey, Ph.D., *Chief, Survey Design Staff*

Andrew A. White, Ph.D., *Chief, Statistical Technology Staff*

Contents

Summary 1

Background 3
Semantic versus episodic memory 3
Telescoping and landmarking 4
Self-concept and event recall 5

Experimental design 7
Issues and hypotheses 7
Design 7
Subjects 8
Methods 8

Results 10
Event recall 10
Dating 16
Health self-assessment 17
Summary 18

Discussion 20

References 22

Appendixes

I. Pilot study 24
II. Sample selection and characteristics 28
III. Experimenter protocol 31
IV. Record matching and scoring 35

List of text tables

A. Experimental procedures outline 9
B. Conditions always regarded as chronic in National Health Interview Survey 10
C. Number of recurring events reported, by experiment 11
D. Pre-postassessment group's recall before the intervention, by event type 11
E. Pre-postassessment group's recall before and after the intervention, by event type 11
F. Pre-postassessment group's incidence of false positives before and after the intervention, by event type 12
G. Recall after the intervention, by subject group and event type 12
H. Overlap between events involving recurring conditions and events involving conditions always regarded as chronic in the National Health Interview Survey (NHIS) 13
J. Recall before and after the intervention, by subject group and event type 14
K. Recall before and after the intervention, by event recency 14
L. Recall before and after the intervention, by subject group and subject's assessment of seriousness 15
M. Recall before and after the intervention by subject group and experimenter's assessment of seriousness 15
N. Mean proportion of medical record events recalled before and after intervention, by subject's reported health condition 16
O. Pre-postassessment group's dating accuracy before the intervention, by event recency 16
P. Dating accuracy before and after the intervention, by subject group and event type 17
Q. Dating accuracy after the intervention, by event type 17

Symbols

- - -	Data not available
. . .	Category not applicable
–	Quantity zero
0.0	Quantity more than zero but less than 0.05
Z	Quantity more than zero but less than 500 where numbers are rounded to thousands
*	Figure does not meet standard of reliability or precision (more than 30-percent relative standard error)
#	Figure suppressed to comply with confidentiality requirements

Autobiographical Memory for Health-Related Events

by Barbara Means, Arti Nigam, and Marlene Zarrow, Human Resources Research Organization; Elizabeth F. Loftus, University of Washington; and Molla S. Donaldson, George Washington University Health Plan

Summary

The National Health Interview Survey (NHIS), administered to 40,000–50,000 households annually, asks respondents to report the number of doctor visits and days of illness they have experienced over the past year. Despite the importance of the accuracy of this information in providing a portrait of the Nation's health needs, survey developers have had little understanding of how respondents answer such questions as "How many times have you seen a doctor for your arthritis since this time a year ago?" Of particular concern has been the reporting of chronic conditions and associated interactions with health care personnel: Survey validation studies have shown that these conditions and incidents are especially likely to be underreported.

Recent developments in the cognitive theory of autobiographical memory provide the basis for a theoretical interpretation of the underreporting of events related to chronic conditions: These conditions are likely to give rise to recurring events, that is, incidents that occur repeatedly in an individual's life and have the same essential characteristics across instances. Psychologists (Neisser, 1986; Linton, 1975) have suggested that such recurring events give rise to a "generic memory" for the group of events and to difficulty in recalling individual instances.

In this study, cognitive theory also provided the basis for the design of a two-part intervention procedure for facilitating the recall of personal events. The personal events consisted of visits made to a health maintenance organization within a 12-month period. The first portion of the intervention, decomposition, was designed to assist subjects in retrieving the individual events (health plan visits) that comprised a generic event memory. Adapting elements of context reinstatement (see Davies, 1986) and cognitive interview techniques (Geiselman et al., 1985), the decomposition procedure asked subjects to think back to the physical setting for the event type and to then try to recall the most recent individual incident of the type. The experimenter used a series of probes concerning aspects of an individual incident—such as getting to the health facility or length of wait—that could set it apart from other similar incidents.

The second portion of the intervention had its theoretical underpinnings in the demonstrated importance of retrieval cues in facilitating recall (Tulving, 1983) and in what is known about the ways in which subjects reconstruct event dates, using stored date tags for a few "landmark" events (Brown, Shevell, and Rips, 1986). This portion of the intervention involved construction of a personal time line, within which the subject tried to fit all remembered health plan visits. The time line was designed both to assist subjects in dating events more accurately and to provide a framework of potential cues that could lead to the recall of additional events.

Subjects drawn from the membership of a health maintenance organization, half of whom reported having a chronic health condition, were asked to describe all of the occasions during the past 12 months on which they had seen a health plan staff member concerning their physical health. A comparison of reported incidents with medical records showed that events of a recurring nature (three or more visits for the same condition during the last 12 months) were recalled more poorly than other events, as predicted.

After this free-recall phase of the experimental interview, the two-part cognitive recall enhancement intervention was administered. The first part of the intervention, decomposition, was used only in those cases where the subject had identified a group of medical visits that he or she said were indistinguishable. After going through the decomposition process to try to recall the most recent such visit, the subject was instructed to apply the same technique to the first visit during the last 12 months and then to all subsequent visits. Finally, all subjects constructed personal time lines, regardless of whether they had gone through the decomposition procedure.

The two-part intervention procedure proved to be effective in enhancing memory for recurring events. After the intervention, recurring events were recalled as well as nonrecurring events. The intervention reduced the proportion of medical record events that subjects failed to recall (false negatives) without increasing the number of reported visits that did not match medical records (false positives). The intervention had a significant positive effect also on the proportion of events for which subjects could provide accurate dates (within 15 days of actual occurrence).

In summary, recall performance before the intervention supported the hypothesis that recurring events lead to the creation of generic memories and to increased difficul-

ties in recalling individual incidents. The success of the intervention lends credence to the proposition that a cognitive model of autobiographical memory and recall is a viable basis for facilitating the recall and dating of personal events.

Background

The National Center for Health Statistics (NCHS), of the U.S. Department of Health and Human Services, administers the National Health Interview Survey (NHIS) to a sample of approximately 40,000–50,000 households of the civilian noninstitutionalized population of the United States each year. The general goal is to describe the social, demographic, and economic aspects of illness, disability, and the use of medical services. The basic items ask individuals to recall injuries and acute conditions, chronic conditions, days of disability, limitations in activity, and hospitalizations for themselves and related household members.

For NHIS purposes, the subject's ability both to recall such health-related incidents and to remember the date on which an incident occurred is important. Some of the survey items ask the respondent to indicate how many incidents of a given type had occurred during the past 2 weeks; other items use a reference period of 3 months; still others use 12 months. In each case, the numbers provided by respondents are weighted to represent an annual rate for the U.S. population as a whole. Thus, the accuracy of respondents' reports of event frequency is a major concern. The incidence rates for various events could be underestimated either if the respondents do not remember relevant incidents which occurred, or if the respondents remember the incidents but incorrectly believe that they happened before the reference period (more than 12 months ago). Incidence rates can be overestimated by the subjects either "remembering" events that did not occur or, if the subjects incorrectly estimate the date of an event, by counting an event as occurring within the reference period when it actually occurred earlier.

The research described in this report investigated people's ability to recall and to date health-related events (such as doctor visits, hospitalizations, and emergency room treatment) that had occurred during the past year of their lives. Conducted by the staff of the Human Resources Research Organization (HumRRO) and the George Washington University (GWU) Health Plan, and by Elizabeth Loftus of the University of Washington, the study drew upon the GWU Health Plan membership as a source of volunteer subjects who were willing to recall health events and to allow their recall to be checked against their actual medical records. In addition to the main study reported here, the project included pilot work, which is described in appendix I. This project was sponsored by the National Center for Health Statistics, with National Science Foundation (NSF) funding, under a continuing program of both in-house and sponsored research designed to test the applicability of cognitive psychology perspectives and research methods to the investigation and reduction of sources of nonsampling error in surveys.

Semantic versus episodic memory

As we began our work in this area, we were interested in the survey research literature finding that underreporting of chronic health conditions (such as rheumatism, allergies, hypertension) is particularly problematic. In an extensive review of the research, Jabine (1985) noted that such underreporting was found when respondents' reports were compared either against medical records or against the results of subsequent physical examinations of survey respondents. It seemed to us that health care events related to chronic conditions might well be recalled differently from events related to acute conditions. We considered the implications of Tulving's (1983) distinction between *semantic memory,* what a person "knows," and *episodic memory,* or memory for particular episodes in the rememberer's life, for understanding recall of the two types of health events. Semantic memory is conceptually structured, whereas episodic memory is thought of as a temporally ordered set of autobiographical events. Thus, an individual uses semantic memory to assert that tuberculosis is a disease of the lungs but calls upon episodic memory to describe a visit to a clinic to get a tuberculosis test. Semantic memory is typically the product of repeated exposures to a piece of information; episodic memory is the trace of a single incident. Tulving argues that access to semantic memory information tends to be immediate (for example, one's response to the question "How many people are in your family?") and resistant to interference from other memory traces, in contrast to access to events in episodic memory ("When was the last time your son saw a doctor?"). Semantic memory reports are typically expressed as "I know that . . ." in contrast to episodic memory reports, "I remember that . . .."

Ostensibly, most questions in a health care survey ask the respondent to draw upon episodic memory, for example, to remember all the occasions on which he or she missed work because of illness during the past year. We suspected, however, that respondents are using much more than pure episodic memory to answer NHIS questions. The

pilot study for this research (described in appendix I) supported this hypothesis. Many responses to health-related questions in the pilot study bore the markers typifying reconstructions or inferences drawn from semantic memory ("I know I usually get a bad cold once or twice a year and that if it's really bad, I take a day off from work"). In addition, when asked to recall all individual visits to doctors, many subjects would recall a typical visit of a particular type (for example, to the allergist or the cardiologist) and report that they had had several such visits but could not remember individual instances. These subjects appeared to be drawing upon semantic memory structures—what cognitive psychologists would call *scripts* or *schemata*—rather than upon episodic memory for individual events.

Schank and Abelson (1977) and Anderson (1984) argue that semantic memory contains scripts or schemata for certain types of recurring experiences. Thus, there is a script or general framework for what happens while dining at a restaurant and one for going to the grocery store, and these will affect the way information is comprehended and remembered (Anderson, Spiro, and Anderson, 1978). Bower, Black, and Turner (1979) found that people have scripts for doctor's office visits. When survey respondents answer questions about particular incidents, we would expect them to recall—or reconstruct—information consistent with the basic script. The pilot study supported this notion: Subjects tended to recall a *typical* visit to a neurologist rather than any *particular* visit. Another script-related phenomenon, found by Bower, Black, and Turner (among others), is that when subjects are asked to recall script-relevant texts, they have a hard time distinguishing what they read (experienced) from things that were not there but are predicted by the script. Extrapolating from these research findings to health surveys, we would predict that if a subject typically misses time from work when she has an asthma attack, she may "remember" missing work when she had an attack last October, even though no work time was actually lost on that occasion. Chronic conditions are more likely than acute ones to be linked to recurring, similar bouts of illness and treatment which can possess the regularity of a script.

Recent work on autobiographical memory has elaborated on Tulving's episodic–semantic distinction and the idea of scripts or schemata for recurring events. A number of researchers have explored the notion that memories for individual life events with shared characteristics can become fused over time, taking on the characteristics of a more global schema. Linton (1986) describes this as a transition from episodic to semantic memory. With time, memories lose their details, and similar events become melded together or are lost. Neisser (1986) thinks in terms of molecular event memories nested within more molar memories; as time passes, the properties that are invariant across the repetitions of an event (the contents of the molar memory) become much more accessible than individual event occurrences. Reiser, Black, and Kalamarides (1986) describe personal memory as consisting of a knowledge structure, corresponding to the schema for a certain type of event, with individual events associated with it. An individual event is marked by indices for those characteristics that distinguish it from the generic or prototype event in the knowledge structure.

This literature offers clues as to why individual health events of a recurring nature should be difficult to remember and why many responses to survey items appear to be drawn from semantic memory. For chronic conditions particularly, an individual will often require repeated interactions with health care providers. Visits are likely to be very similar—to involve the same health care providers, occur at the same location, and consist of the same actions (for example, receiving an allergy shot or getting blood pressure checked). Over time, individual visits to the health facility for such a condition would tend to blend together into a "script" and the script would be more detailed and specialized than the general script for doctor visits. A specialized script, such as that for seeing an allergist at the GWU Health Plan, contains concrete detail ("I go to the third floor, give my name to the nurse, and then Ms. Jones calls me in for my allergy shot.") than the general doctor visit script ("I go to the doctor's office, wait, and then see the doctor.").

We can think about the recall of visits for chronic and acute conditions in terms of Reiser, Black, and Kalamarides' (1986) model: Associated with a script will be relevant individual events which are indexed in terms of those aspects of the individual event that deviate from, or go beyond, that which is specified in the script. These markers for script deviations will be the key to retrieval of individual events. In the case of chronic conditions, the specialized script will contain so many specific elements of each health plan visit that there will be few distinguishing elements that can be used to cue recall of individual incidents.

In addition, script intrusions should be harder to identify in the case of visits of a recurring type. Because the script is specialized, script intrusions are not different in surface form from episodic memory of particular events. In contrast, health plan visits related to acute conditions tend to be unpredicted and unique. Such visits may be linked to a less fully specified "doctor's office visit" script ("I see a doctor and he examines me"), in which case script intrusions should not have the level of detail associated with memory of individual events.

In summary, for random, nonrecurring health care events, such as the treatment of an injury, we would expect recall of individual event traces. For recurring health care events of the same type, we would expect subjects to draw heavily upon the specialized script for that type of visit and to have trouble recalling distinguishing details of individual incidents.

Telescoping and landmarking

The accuracy of NHIS results depends not only on the respondents' ability to retrieve health care events from

memory, but also on their ability to accurately date those incidents. Survey researchers have long noted a tendency for respondents to transport events that occurred before the reference period forward in time, a phenomenon called *forward telescoping* (Sudman and Bradburn, 1973). Instances of dating an event as older than it really is, *backward-telescoping,* occur as well, although less frequently than forward-telescoping.

Cognitive research on the ability to remember dates for public or private real-world events has concluded that date tags are not stored with most events. Rather, people appear to have a few privileged, or landmark events, for which they know dates, and to reason about dates for other events based on relationships with the landmark event. Linton (1975) describes the use of such "markers" in her study of her own memory for autobiographical events in her life. Brown, Shevell, and Rips (1986) report that their subjects used the same kind of process in dating public events, often using landmark events from their personal lives that could be associated with the public event that they were trying to date. In a review of think aloud protocols collected during the pilot-study phase of our research, the subjects estimated the dates for health events by using a variety of landmark events, including holidays, birthdays, weddings, and major health events. There were several instances of apparent storage of a date tag with an important health care event—such as a birth or open-heart surgery. The date for this event appeared to be accessed directly, and the event served as a natural cue and dating landmark for estimating dates of other health care events. Because a series of doctor visits often precedes a major health event like a birth or operation, and followup visits occur at regular intervals, major health care events can serve as very effective temporal markers.

The way in which people use landmark events to reason about the recency of other events that they remember has implications for the improvement of survey methods. Baddeley (1979) found that subjects who spontaneously use personal landmarks tend to be more accurate in their recollections than people who do not. Loftus and Marburger (1983) found that retrospective autobiographical events were better dated, with less forward-telescoping, if subjects were instructed to use either personal or public events as landmarks ("Since the eruption of Mount Saint Helens, has anyone beaten you up?"). Thus, the research on how people date events in autobiographical memory, like the discussion of memory for recurring events, implies that responses to survey items will depend heavily on inferential processes. The research on dating suggests further that survey techniques that help the respondents capitalize on landmark events would improve the subjects' ability to discriminate events that fall within the reference period from those outside the period.

Work by Brown, Rips, and Shevell (1985) suggests that there may be differences between different types of events in the extent to which they are susceptible to forward-telescoping. In their study, subjects estimated which of a pair of public events had occurred more recently. They found that events that are more accessible, like Hinckley's attempt to assassinate President Reagan, are displaced toward the present more than similar but less accessible events, like the assassination attempt against the Pope. This accessibility hypothesis provides an interpretation for the phenomenon of forward-telescoping and suggests that the phenomenon would be more pronounced for more salient health care events.

Although the findings reported by Brown, Rips, and Shevell (1985) are provocative, we emphasize caution in extrapolating to NHIS. The events for which these subjects gave recency estimates were public events, not personal ones, and had occurred as much as 10 years before the experiment. For events in the respondent's life that occurred within the timeframe of interest for NHIS (that is, 12 months), the mechanisms for dating events and the effects of event seriousness may be quite different.

An alternative hypothesis concerning the relationship between an event's seriousness and dating accuracy would be that serious events are more likely than minor events either to be landmark events themselves or to have associations with landmark events that will permit inferences concerning their date of occurrence. Thus, we would expect serious events to be dated more accurately and to show less deterioration in dating accuracy over time than minor events. In the pilot study, the experimenters classified health events as serious or minor. A number of serious events were dated precisely and appeared to function as landmark events, as described above. Dating accuracy overall was better for those events classified as serious by the experimenters, supporting the notion that at least over a period of a year, serious events are dated more accurately than minor ones. However, the categorization of events into the two seriousness categories was performed post hoc by the researchers; what we really want to know is how the seriousness or salience of the event for the subject affects the perceived recency of the event. In the present study, subjects rated the seriousness of all the events they had recalled at the conclusion of the experimental session. Thus, we could explore both the accuracy of dating for health events over different time lags and the impact of perceived event seriousness on recency estimates. If the accessibility hypothesis of Brown, Rips, and Shevell (1985) pertains to relatively recent personal events, there should be a tendency toward forward-telescoping that is more marked among those events regarded as serious. If, on the other hand, serious events are more likely to have stored date tags or a stable set of associations with other events from which dating inferences can be drawn, dating should be more accurate for the events that subjects regard as serious.

Self-concept and event recall

A final hypothesis addressed by this study concerns the effect of the respondent's self-perception of his or her health on the accuracy of reporting and the retrieval strategies used. Jabine (1985) reports that in a

Kaiser-Permanente study (NCHS, 1967, 1973) the extent of underreporting of chronic conditions was strongly related to the respondent's overall perception of his or her general health. Those who viewed themselves as very healthy were most likely to fail to report conditions found in their medical records. From the standpoint of schema theory, this makes sense. The respondent starts with the overall concept "there is nothing wrong with me." Given specific conditions or symptoms on the checklist, the respondent does not do an exhaustive memory search of all health incidents, seeking an experience with each one. A more likely retrieval strategy used for a particular respiratory condition, for example, would be accessing a schema regarding one's state of health and noting that there are no respiratory problems within that schema. Barclay's (1986) argument, that the same self-schemata control both attention to self-referenced material and the reconstruction processes that constitute most of autobiographical memory, is consistent with this line of reasoning. Thus, we wished to explore the hypothesis that individuals with a very positive concept of their own overall health would be more likely than other subjects to underreport health visits on their medical records.

Experimental design

Issues and hypotheses

This study was designed to investigate several factors that could affect the accuracy of recall and dating for health-related incidents. We hypothesized that the way in which such incidents would be recalled, their frequency estimated, and event dates remembered would vary depending on the nature of the events. A series of very similar events, such as repeated doctor visits to obtain allergy shots, was hypothesized to give rise to generic-event memories. Individual instances of generic-event memories should be more difficult to recall than nonrecurring events.

Both our pilot study and the work of Brown, Rips, and Shevell (1985) suggested that another factor affecting recall would be the seriousness of the health-related incident. However, these two lines of work suggested different effects. One possibility is that the probability of recall is higher and the accuracy of date estimates is better for events that the subject perceives as serious. The alternative prediction, based on the accessibility hypothesis of Brown, Rips, and Shevell (1985), is that serious events would be dated less well than minor events because they would be more susceptible to forward-telescoping.

The final issue addressed by this study concerned the subject's self-concept regarding his or her own health. The hypothesis was that respondents with extremely positive assessments of their own health would tend to underreport utilization of health care facilities.

The present study expanded upon the pilot study by introducing an intervention procedure designed to positively influence the way in which health-related incidents are recalled and dates remembered or estimated. In the pilot study, many subjects recalled groups of visits for the same condition, without remembering details specific to any one visit or providing a judgment regarding the total number of such visits. When probed, these subjects tended to respond that these visits could not be remembered individually because they were "all the same." In the present study, the first part of the intervention was designed to deal with such situations. This part of the intervention—decomposition—involved asking the subject first to remember some event-specific details about the last visit of the group. (The last visit was used to prompt episodic recall because it would have the greatest recency and, all things being equal, should be the easiest of the visits to remember in detail.) This procedure was similar to the "guided memory" technique that has been used in studies of eyewitness testimony (Davies, 1986; Smith, 1979).

The experimenter prompted the subject to visualize the room in which the visit took place and then to try to recall events surrounding the visit, such as taking off from work, getting to the health plan office, and waiting for the doctor. After eliciting an episodic description of the last visit, the experimenter then asked the subject to remember specific details about the first such visit, then the first visit during the recall period, and all subsequent visits.

The decomposition procedure was used only for subjects who recalled some set of multiple events as a group, without providing recollections of individual visits when the experimenter requested them. It was expected that most of these "grouped" visits would involve chronic conditions, perhaps accounting for some of the underreporting of chronic conditions reported in the survey literature (Jabine, 1985).

The second portion of the intervention, which was used with all subjects subsequent to their first attempt at recall, involved construction of a personal time line for the 18 months preceding the month of the interview. This time line was intended to stimulate autobiographical memory of important landmark events during this period; these events could then be used as a structure within which the subject could try to recall and date the health-related incidents of interest to NCHS. Subjects had been observed to use personal landmarks to trigger recall and estimate dates in our pilot study and also in other work sponsored by NCHS with National Science Foundation funding (NCHS, 1989). Our time line intervention systematized this process, prompting all subjects to identify and date such landmark events (for example, birthdays, holidays, promotions, vacations) and then to use them as cues for remembering and dating health-related incidents. We hypothesized that the construction and use of a personal time line would lead to the generation of additional recall cues that would trigger recall of additional health-related incidents, as well as an improvement in ability to accurately date the incidents already remembered.

Design

To obtain the maximum amount of information from a relatively small sample of subjects, we presented the subjects with a two-part recall task: (1) The experimenter

instructed the subjects to recall each time during the past 12 months that they had seen someone at the GWU Health Plan about their health. This was the "free recall" task. (2) The experimenter helped the subjects to construct a personal time line and try to place all health-related visits within this context ("time line" task). Those subjects requiring the decomposition portion of the intervention received it after the free recall but before the time line.

Assessment of recall accuracy requires matching events in recall protocols to those in the subjects' medical records. To make this matching possible, we had to get information from the subjects' recall protocols concerning not only the nature of a health plan visit or hospitalization but also the date of the visit, health care provider, treatment, and so on. This information was requested and recorded on "information cards." To estimate the accuracy of recall before the intervention, we had the subjects provide the various types of information needed in verifying recall against medical records for all remembered events just after free recall.

To score recall accuracy after the intervention, we had all subjects complete information cards at the end of the session after filling in the time line. Under certain conditions, repeated testing of recall can itself improve performance (Roediger and Payne, 1982); therefore, we wanted to rule out the possibility that any improvements in recall and dating accuracy following the intervention stemmed from two opportunities to provide the data for the information cards, rather than from the intervention itself. Accordingly, we had half of the subjects provide responses for information cards after free recall and at the end of the session, whereas the other subjects filled out cards only at the end of the session. The first group will be referred to as the *pre-postassessment* group and the second, as the *postassessment-only* group.

The information cards filled out by the pre-postassessment group immediately after free recall were used to estimate recall and dating accuracy before the intervention. Accuracy after the intervention was measured from the information card data provided by the postassessment-only group at the end of the session. By comparing recall and dating accuracies based on final information cards for the pre-postassessment group and the postassessment-only group we could assess whether the extra opportunity to fill out the information cards had an effect on recall and dating accuracy. If not, final recall and dating data for the two assessment groups could be combined for comparisons with pre-postassessment group recall and dating accuracy after free recall.

Limitations on the number of subjects who could be tested precluded use of a third group for which information cards would be obtained after the decomposition portion of the intervention. Hence, this design permits an assessment of the efficiency of the intervention as a whole but does not enable a strict test of the utility of the individual parts of the intervention.

Subjects

Subjects were selected from the membership of the George Washington University (GWU) Health Plan, using the procedures for identifying eligible participants described in appendix II. Responses to the plan's annual membership survey were used to select potential respondents from plan members whose medical experiences over the past year would provide a reasonably demanding memory task: Only those members who reported having had four or more health plan visits over the last year were identified. From this group we drew half of our potential subjects from those who said they had some kind of chronic health problem.

In all, 51 plan members were sent a letter and subsequently contacted by telephone and asked to participate. The purpose of the project, including its voluntary nature and the requirement to check what the participants recalled against their medical records, was explained. Potential subjects were offered $20 in compensation for their time. Those subjects who agreed to participate were asked to rate their overall health as excellent, very good, good, fair, or poor. The order in which response options were presented was alternated across subjects.

Of the 51 potential subjects contacted, 43, or 84 percent, agreed to participate. One of these subjects later canceled and did not reschedule her appointment, and one had to leave her experimental session before completion. Data from an additional subject were discarded because of experimenter error in conducting the session, leaving a total of 40 subjects. Of these, 20 were men and 20 were women, all within the ages of 24 and 64. Twenty-four had a college degree or more education. The mean number of health plan visits per subject for the reference period was 7.75, with a range of 3 to 16.

Methods

The subjects participated in individual experimental sessions, lasting between 1 and 1½ hr. Sessions were held in the subjects' homes or at the GWU Health Plan or HumRRO offices, depending on the subject's preference. The subjects first were asked how many times they had seen a medical doctor, nurse practitioner, or physician's assistant about their health over the last 12 months. The experimenter then led them through guided practice in giving verbal protocols. Following this practice, the subjects were asked to think aloud as they recalled each incident, describing as much about it as they could. During this free recall, the experimenter started an information card for each event, writing a few words on a blank card to represent each visit.

When the subjects indicated that they had recalled all relevant events, they were asked to give a new assessment of the number of health care visits they had had in the past year.

To allow for accurate matching between recalled events and medical record entries, pre-postassessment group subjects were asked to provide all the types of information on each visit that could be used in identifying particular events (reason for visit, date of visit, treatment, medical tests, visit type, type of medical professional seen and associated medical team, name of person seen). This information was recorded on the information card for each event recalled by a pre-postassessment group subject during free recall.

After completing the information cards (for the pre-postassessment group) or immediately after free recall (postassessment-only group), the experimenter determined whether the subject had reported multiple visits of the same type without recalling each individually (for example, "I went to the allergist for shots, probably 12 times"). Those subjects who had demonstrated such generic event memories were then given the decomposition portion of the intervention. The experimenter used a series of probes to try to elicit episodic memories for these events. The experimenter explained that events that are all similar are difficult to remember individually but that she would like him or her to try to isolate each specific visit. The subject was directed to start with the last visit of this type and to try to go back to the physical context of the visit and then remember some differentiating detail from that visit (for example, difficulty getting to the office, disruption of other plans, an unusually long or unusually short wait). Next, the same technique was tried for the first of these visits and then for each of these visits in the reference period. When subjects could recall no more individual events of this type, the experimenter invited them to revise their estimate of the total number of visits for the year if they so desired.

In the second part of the intervention, the experimenter worked with all subjects to construct an individual time line for the 18 months preceding the interview. (The 6 months before the 12-month reference period were added to help the subjects identify events that had been mistakenly forward-telescoped into the reference period.) Each subject was given a long sheet of paper containing a printed time line marked off in months with holidays indicated (as illustrated in appendix III). The subject was then asked to write in personal landmarks, such as "my birthday," "the day I started my new job," and so on. The experimenter encouraged the subject to think of at least one landmark event for each month. Once the landmarks were inserted on the time line, the experimenter asked the subject to insert all health care visits at appropriate points. The subjects were urged to try to remember any visits they had not thought of previously. After completing this process, the experimenter counted the number of visits in the 12-month reference period. The subject was asked to assess the accuracy of this number and was invited to revise his or her estimate of the number of health care visits in the year.

The experimenter made information cards for any additional visit recalled during decomposition or the time line. All subjects either completed (in the case of the postassessment-only group) or amended (pre-postassessment group) the information cards after completing the time line. The subjects then rated the seriousness of the health condition related to each visit they had remembered, using a five-point scale ("very minor, minor, moderate, serious, very serious") displayed on a sheet of paper.

The experimental procedures described above are outlined in table A. The experimenter's protocol (excluding think-aloud probes, which depended on the subject's responses) is contained in appendix III.

At the conclusion of the session, the subjects were asked to rate their health as excellent, very good, good, fair, or poor. The response choices were presented either with "excellent" first or with "poor" first, in the same order used by the health plan representative who had set up the interview for that subject. Thus, each subject responded to the same health assessment question twice: Once over the telephone, some days prior to the interview when asked by the health plan representative, and once when asked by the experimenter at the conclusion of the session.

Table A. Experimental procedures outline

1. Health assessment (during telephone scheduling)
2. Initial estimate of number of health care visits in last 12 months
3. Think aloud practice
4. Free recall of each incident (think aloud)
5. Second estimate of number of health care visits
6. Provision of information for medical record verification (pre-postassessment-group subjects only)
7. Decomposition of generic memories for grouped events, if required
8. Estimate of number of health care visits (decomposition intervention subjects only)
9. Development of personal time line and insertion of health care visits into time line
10. Final estimate of number of visits in last 12 months
11. Provision or revision of information for medical record verification
12. Rating of seriousness of each event
13. Health assessment (at end of session)
14. Decision process for serious–minor classification

Results

Event recall

Recurring events

The intervention procedure used in this study was designed to enhance memory for similar, recurring events on the assumption that such events tend to give rise to generic memories, with a concomitant difficulty in retrieving individual, distinctive instances from the event group. From the medical record abstracts for the subjects, we were able to identify the recurring events in the medical history of each study participant. "Recurring events" are defined as those health plan visits for a condition that had led to three or more visits during the last 12 months. The set of three or more medical record events related to a single condition is referred to as an "event group." Our assumption was that such event groups are likely to be represented in memory as generic memory traces, as discussed earlier.

Medical records showed a total of 40 event groups for our subjects, with the groups containing a total of 168 events (a mean of 4.20 events per group). Fourteen of these event groups involved conditions that are always considered chronic by NCHS (see table B), but other event groups involved conditions not usually considered chronic (for example, ear infection), or even repeated followup visits for an injury. We decided to focus on the recurring versus nonrecurring distinction rather than on the distinction between chronic and nonchronic events in our primary set of analyses because, from a psychological perspective, it is the predictability and frequency of an event type, rather than the nature of the medical diagnosis category per se, that is relevant to the development of generic memories.

Table B. Conditions always regarded as chronic in National Health Interview Survey

Tuberculosis
Neoplasms (benign and malignant)
Diseases of the thyroid gland
Diabetes
Gout
Psychoses and certain other mental disorders[1]
Multiple sclerosis and certain other diseases of the central nervous system
Certain diseases and conditions of the eye
Certain diseases of the circulatory system (includes rheumatic fever, hypertension, stroke, and all heart conditions)
Emphysema, asthma, hay fever, and bronchiectasis
Ulcers and certain other diseases of the esophagus, stomach, and duodenum
Hernia of abdominal cavity (includes rupture)
Gastroenteritis and colitis (with exceptions)
Calculus of kidney, ureter, and other parts of the urinary system
Diseases of the prostate
Chronic cystic diseases of the breast
Eczema and certain other dermatitis
Arthritis and rheumatism
Cyst of the bone (except jaw)
All congenital anomalies

[1]Psychiatric visits were not included in the Human Resources Research Organization–George Washington University Health Plan study.

Source: National Center for Health Statistics, B. Bloom. 1982. Current estimates from the Health Interview Survey, United States, 1981. Vital and Health Statistics. Series 10, No. 141, p. 49. DHHS Pub. No. (PHS) 82–1569. Public Health Service. Washington: U.S. Government Printing Office.

Table C provides a summary of the subjects' reporting of events from these event groups during the course of the experiment. These data show the first mention of an event, without regard to the completeness or accuracy of the event description. During free recall, when asked to recall all health plan visits during the past year while thinking aloud, the subjects mentioned 77 events from the 40 groups of recurring events on their records, an average of 1.93 events per group. For seven event groups, no individual incidents were reported during free recall. For 15 of the event groups, the subjects explicitly said that there were multiple visits of the same type but that they could not remember how many. These event groups were flagged for the decomposition procedure. For these groups, decomposition resulted in reports of an additional 35 events, a mean increment of 2.33 events for each group. During the time line portion of the intervention, an additional 18 recurring events were reported, 8 from event groups that had been decomposed and 10 from other recurring event groups. At the conclusion of the intervention, the mean number of events reported for groups that had been decomposed was 4.00 (88 percent of the actual), whereas that for groups that had not been decomposed was 2.80 (70 percent of the actual), with a combined mean of 3.25.

Although it is clear that the intervention increased the mean number of events reported for the 40 recurring event groups (3.25 vs. 1.93, $t(39) = 4.61, p < 0.001$), it cannot be assumed that increases in the number of events reported necessarily mean increases in recall accuracy (as measured by date, and so forth). To assess the effect of the intervention upon recall accuracy, we had to compare the subjects' reports to their medical record entries and establish a baseline recall accuracy level after free recall. This was done by comparing medical record information with the data provided on information cards filled out immediately after free recall by the pre-postassessment group of subjects.

Table C. Number of recurring events reported, by experiment phases

Decomposition status	Number of event groups	Actual mean number of events per group	Mean number of events per group: Cumulative	Reported in free recall[1]	First reported in decomposition	First reported in time line
All recurring events	40	[2]4.20	3.25	1.93	...	0.45
Flagged for decomposition	15	4.53	4.00	1.13	2.33	0.53
Not flagged for decomposition	25	4.00	2.80	2.40	...	0.40

[1]Events reported include false positives as well as accurately recalled events. Subjects who describe a recurring event from a particular group, without identifying more than 1 individual instance, were scored as reporting one event from the group. (They had been asked earlier to give a number.)
[2]Some event groups overlap. Although there are a total of 153 events classified as "recurring" for all subjects, the total of the group sizes for the 40 event groups is 168.

Table D. Pre-postassessment group's recall before the intervention, by event type

Event type	Number of events on medical record (1)	Matches: Number (2)	Matches: Number per record event (3)	False negatives: Number (4)	False negatives: Number per record event (5)	False positives: Number (6)	False positives: Number per record event (7)	Gross error rate (8)	Net error rate[1,2] (9)
All events	139	57	.41	82	.59	39	.28	.87	.31
Recurring events	81	26	.32	55	.68	14	.17	.85	.51
Nonrecurring events	58	31	.53	27	.47	25	.43	.90	.03

[1]The gross error rate equals the sum of column 5 and column 7.
[2]The net error rate equals the difference between column 5 and column 7.

Procedures used in matching events described on information cards to medical record entries, and criteria for scoring matches, are described in detail in appendix IV. Three recall measures were computed. Recalled events that could be matched to events on medical records were counted as *matches.* Events on medical records that subjects did not recall were termed *false negatives.* Recalled events for which there was no corresponding medical record entry were termed *false positives.*

For the 81 recurring events in 20 event groups on the medical records of the pre-postassessment group, 26 events or 32 percent were described accurately enough during free recall to be matched to a medical record entry. This rate is lower than the recall rate for nonrecurring events, which was 53 percent, $Z = 2.52, p < .05$, as predicted by the model of autobiographical memory described earlier. These free recall data for the pre-postassessment group are shown in table D.

Recall performance after the intervention for recurring and nonrecurring events in the medical histories of pre-postassessment group subjects is shown in table E. The intervention led to improved recall for recurring events: After completion of the time line, the pre-postassessment group recalled 67 percent of the recurring events on their medical records compared with just 32 percent after free recall (Cochran's Q (1, $N = 81$) = 24.50, $p < .01$). This improvement was more marked for those recurring events that had been the target of the decomposition procedure. For subjects in the pre-postassessment group, 7 of the 20 event groups underwent decomposition. For the 32 events contained in the 7 recurring event groups, recall of record events rose from a free recall level of 16 percent to an after-intervention level of 78 percent (Cochran's Q (1, $N = 32$) = 6.67, $p < .01$). The remaining 49 events were in the 13 event groups that were not flagged for decomposition, because the subject's free recall had failed to indicate that there were multiple unreported events. The improvement for these recurring events was still notable: It went from a recall level of 43 percent after free recall to one of 59 percent after the time line (Cochran's Q (1, $N = 49$) = 8.00, $p < .01$). Thus, the time line alone produced significant improvement in recall of recurring events.

The intervention was designed with the difficulties of recalling recurring events kept in mind. Not surprisingly, its effect was less potent in enhancing memory for nonrecurring events. For nonrecurring events, thepre-postassessment group's recall improved from the free recall level of 53 percent to 59 percent after the time line. This improvement was not significant (Cochran's Q (1, $N = 58$) = 1.00, $p > .10$).

Table E. Pre-postassessment group's recall before and after the intervention, by event type

Event type	Number of events on medical record	Medical record events recalled: Before intervention	After intervention
		Percent	
All events	139	41	63
Recurring events			
Total	81	32	67
Flagged for decomposition	32	16	78
Not flagged for decomposition	49	43	59
Nonrecurring events	58	53	59

NOTE: Based on data for the 18 pre-postassessment group subjects.

Table F. Pre-postassessment group's incidence of false positives before and after the intervention, by event type

Event type	Number of events on medical record	False positives before intervention		False positives after intervention	
		Number	Percent	Number	Percent
All events	139	39	28	33	24
Recurring events:					
Flagged for decomposition	32	1	3	3	9
Not flagged for decomposition	49	13	27	8	16
All recurring events	81	14	17	11	14
All nonrecurring events	58	25	43	22	38

Strictly speaking, the contribution of the decomposition portion of the intervention to recall performance cannot be separated from that of the time line as the information needed to assess recall level was collected (that is, information cards were filled out) at the end of the time line only. However, the considerable number of new events generated during decomposition (see table C), and the fact that after-intervention recall for recurring events that underwent decomposition (78 percent) tended to be higher than that for recurring events that did not (59 percent) ($Z = 1.77$, $p < .10$), leads to the inference that decomposition contributed to the overall success of the intervention. A major difficulty with the decomposition procedure, however, is that the experimenter lacks reliable information about when to apply it. Subjects' responses during free recall identified only 15 of the 40 event groupings that could, theoretically, have profited from this portion of the intervention.

It is important to note that the improvement in the proportion of medical record events recalled, as brought about by the intervention, did not come at the expense of increased rates of false positives. The number of false positives after free recall and after the intervention is shown in table F. For the pre-postassessment group, there were 39 false positives before the intervention and 33 afterward (Cochran's Q $(1, N = 49) = 1.38, p > .10$).

Contrary to expectation, the ratio of false positives to events on record among pre-postassessment group subjects was higher for nonrecurring events—.43 before the intervention and .38 after—than for recurring events—.17 before the intervention and .14 after. The difference between event types was significant both before, $Z = 3.34$, $p < 0.001$, and after, $Z = 3.33$, $p < .001$, the intervention. (The effect was similar for the postassessment-only group, with false positive rates of .42 for nonrecurring and .25 for recurring events: $Z = 2.36$, $p < .05$.) The difference in false positive rates for recurring events flagged for decomposition (3 percent) and those that were not flagged (27 percent) suggests that subjects may avoid false positives for recurring events by not attempting to recall individual incidents.

Tables D and G, which summarize recall performance before and after the intervention, respectively, also show the gross and net error rates before and after the intervention. "Gross error rate" is defined as the sum of false negatives and false positives divided by the number of events on the medical records. "Net error rate" is the difference between false negatives and false positives divided by the number of events on the medical records. Net error rate is a type of sensitivity measure, reflecting the subject's tradeoff between the risk of a false positive

Table G. Recall after the intervention, by subject group and event type

Subject group and event type	Number of events on medical record (1)	Matches		False negatives		False positives		Gross error rate[1] (8)	Net error rate[2] (9)
		Number (2)	Number per record event (3)	Number (4)	Number per record event (5)	Number (6)	Number per record event (7)		
Pre-postassessment group:									
All events	139	88	.63	51	.37	33	.24	.60	.13
Recurring events	81	54	.67	27	.33	11	.14	.47	.20
Nonrecurring events	58	34	.59	24	.41	22	.38	.79	.03
Postassessment-only group:									
All events	171	97	.57	74	.43	60	.35	.78	.08
Recurring events	72	39	.54	33	.46	18	.25	.71	.21
Nonrecurring events	99	58	.59	41	.41	42	.42	.84	-.01
Total sample:									
All events	310	185	.60	125	.40	93	.30	.70	.10
Recurring events	153	93	.61	60	.39	29	.19	.58	.20
Nonrecurring events	157	92	.59	65	.41	64	.41	.82	.01

[1]The gross error rate equals the sum of column 5 and column 7.
[2]The net error rate equals the difference between column 5 and column 7.

response and that of a false negative. From the survey analyst's viewpoint, a net error rate of zero is desired.

Table D shows that before the intervention, gross error rates were very similar for recurring and nonrecurring events (.85 and .90) but net error rates were quite different (.51 vs. .03, $Z = 5.55, p < .0001$), reflecting the higher false negative rate and lower false positive rate for recurring events. The net error rates in table G indicate that despite the reduction in false negative rates after the intervention, a tendency toward underreporting of events from the recurring event groups persists (net error = .20).

Table G also contains data pertinent to comparing the recall performance after the intervention for the two assessment groups. The percentage of medical record events recalled after the intervention by the two groups was quite similar overall, 63 and 57 percent ($Z = 1.18, p < .23$). The error patterns specifically for nonrecurring events were also quite similar for the two groups: Gross error rates, for example, were .79 and .84 ($Z = .72, p < .48$). The two assessment groups differed, however, in terms of errors connected with recurring events. The gross error rate for recurring events among pre-postassessment subjects was .47, compared with .71 for postassessment-only subjects ($Z = 2.99, p < .01$). This difference reflected a trend toward both a lower false negative rate (.33 compared with .46 for the postassessment-only group, $Z = 1.58, p < .12$) and a lower false positive rate (.14 compared with .25, $Z = 1.80, p < .10$) in the pre-postassessment group; this suggests that completing information cards before the intervention did have some positive effect on memory for recurring events. It seems reasonable to suppose that the request to provide event-specific information on the card helped the subjects to organize their memories for the class of events and led them to begin to decompose their generic memories for these event groups.

Event chronicity

The above results contrast recurring and nonrecurring events, objectively defined on the basis of medical record entries. The survey researcher, however, does not have access to such records and is concerned with medical definitions of chronicity, relying heavily on prespecified lists of conditions that are regarded as chronic. To ascertain whether difficulty in remembering recurring events could account for the greater incidence of underreporting for chronic conditions in health surveys, we classified each event with regard to chronicity as commonly defined by NHIS. We sought to ascertain the extent to which our intervention would prove effective in enhancing the reports of medical events related to chronic conditions per se. Each event was evaluated to determine whether it concerned a condition that is always considered chronic by NHIS (these conditions are shown in table B). Table H shows the overlap between the chronic- and recurring-event classifications.

Table H. Overlap between events involving recurring conditions and events involving conditions always regarded as chronic in the National Health Interview Survey (NHIS)

Event type	Total	Number of events: On NHIS chronic condition list	Number of events: Not on NHIS chronic condition list
All events on medical record. . . .	310	58	252
Recurring events on medical record	153	49	104
Nonrecurring events on medical record (except routine events) . .	130	9	121
Routine events on medical record	27	0	27

As would be expected, most of the events related to conditions that are always considered chronic (49 of 58 chronic events) had been classified as recurring on the basis of the medical record data. However, many events associated with conditions not on the always-chronic list also met our criterion for classification as recurring. For the most part, then, chronic events were a subset of the larger category of 104 recurring medical record events, used in the analyses reported above.

The intervention did produce significant improvement in recall of events related to chronic conditions (table J). For the pre-postassessment group, the match rose from 27 percent before the intervention to 78 percent afterward (Cochran's Q 1, $N = 37) = 19.00, p < .01$). The percent of chronic events recalled by the postassessment-only group, 62 percent, was also significantly greater than that recalled by the pre-postassessment group before the intervention ($Z = 2.64, p < .01$).

Table J shows also that the intervention was quite beneficial for recurring events, whether or not they involved conditions that are always chronic. Recall for the pre-postassessment group improved from 21 to 76 percent for chronic-recurring and from 40 to 60 percent for nonchronic-recurring events (Cochran's Q (1, $N = 33) = 18.00$, $p < .01$ and Q (1, $N = 48) = 7.14$, $p < .01$).

The trend toward more complete recall after the intervention for chronic- than for nonchronic-recurring events in this group (76 percent compared with 60 percent, $Z = 1.74$, $p < .10$) was not found in the postassessment-only group, within which recall rates were nearly identical (56 percent and 54 percent) for chronic- and nonchronic-recurring events.

In contrast to the striking improvement in recall of both chronic-recurring and nonchronic-recurring events, those events that were neither chronic nor recurring were not recalled better after the intervention. Recall rates for these events were 52 percent before and 56 percent after the intervention ($Z = 0.39$, $p < .69$). This category did undergo a significant reduction of gross error, however, from .94 to .81 ($Z = 2.07, p < .05$).

In conclusion, then, the intervention made a significant improvement in recall for both chronic events, as defined in NHIS, and recurring events, as defined on the basis of medical records.

Table J. Recall before and after the intervention, by subject group and event type

Intervention status, subject group, and event type	Number of events on medical record (1)	Matches: Number (2)	Matches: Number per record event (3)	False negatives: Number (4)	False negatives: Number per record event (5)	False positives: Number (6)	False positives: Number per record event (7)	Gross error rate[1] (8)
Before intervention								
Pre-postassessment group:								
All events	139	57	0.41	82	0.59	39	0.28	0.87
Chronic	37	10	0.27	27	0.73	1	0.03	0.76
Recurring	33	7	0.21	26	0.79	1	0.03	0.82
Nonrecurring	*4	*3	*0.75	*1	*0.25	–	*0.00	*1.00
Nonchronic	102	47	0.46	55	0.54	38	0.37	0.91
Recurring	48	19	0.40	29	0.60	13	0.27	0.88
Nonrecurring	54	28	0.52	26	0.48	25	0.46	0.94
After intervention								
Pre-postassessment group:								
All events	139	88	0.63	51	0.37	33	0.24	0.60
Chronic	37	29	0.78	8	0.22	2	0.05	0.27
Recurring	33	25	0.76	8	0.24	–	–	0.24
Nonrecurring	*4	*4	*1.00	–	–	*2	*0.50	*0.50
Nonchronic	102	59	0.58	43	0.42	31	0.30	0.73
Recurring	48	29	0.60	19	0.40	11	0.23	0.63
Nonrecurring	54	30	0.56	24	0.44	20	0.37	0.81
Postassessment-only group:								
All events	171	97	0.57	74	0.43	60	0.35	0.78
Chronic	21	13	0.62	8	0.38	12	0.57	0.95
Recurring	16	9	0.56	7	0.44	5	0.31	0.75
Nonrecurring	*5	*4	*0.80	*1	*0.20	*7	*1.40	*1.60
Nonchronic	150	84	0.56	66	0.44	48	0.32	0.76
Recurring	56	30	0.54	26	0.46	13	0.23	0.70
Nonrecurring	94	54	0.57	40	0.43	35	0.37	0.80

[1]Gross error rate is the sum of column 5 and column 7.

Table K. Recall before and after the intervention, by event recency

Time between event and interview	Before intervention (pre-postassessment group): Number of events on medical record	Before intervention (pre-postassessment group): Percent recalled	After intervention (total sample): Number of events of medical record	After intervention (total sample): Percent recalled
All events	139	41	[1]309	60
0–1 month	14	71	35	83
2–3 months	16	69	42	76
4–8 months	59	39	126	65
9–12 months	50	26	106	39

[1]1 event was dropped because correct date was unknown.

Event recency

Another factor related to recall performance was the recency of the event. Table K shows the effects of the intervention for events of various recencies. Although events that were no more than a month old were well remembered even before the intervention (71 percent), recall before the intervention dropped off rapidly for older events—only 26 percent of events 9–12 months old were recalled. The intervention had the biggest impact on events 4–8 months old, raising the mean recall level from 39 to 65 percent ($Z = 3.34, p < .001$).

Event seriousness

We had predicted that recall and dating accuracy would be influenced by the seriousness of the event as perceived by the individual subject. To measure these effects, subjects were asked to rate the seriousness of each event they recalled on a five-point scale. Because seriousness ratings clustered around the midpoint of the scale, the events that subjects rated as "very serious" and "serious" were combined as were those regarded as "minor" and "very minor." Because subjects could not provide seriousness ratings for events that they failed to recall but which were on their medical records (false negatives), we could not use proportion of medical record events recalled as a criterion when assessing the effects of subjective perception of event seriousness. Instead, we employed an alternate measure—the percent of recalled events that had a match on the medical record (matches/matches + false positives). Table L contains these data. In terms of recall after the intervention, 60 percent of the 77 recalled events that subjects rated as "serious" or "very serious" were matches to medical record events, compared with 70 percent of 101 "in between" events, and 69 percent of 98 "minor" and "very minor" events.

Table L. Recall before and after the intervention, by subject group and subject's assessment of seriousness

Intervention status, subject group, and seriousness rating	Events recalled	Events matched to medical record	False postives: Total	False postives: Forward-telescoping	False postives: Other	Percent of events matched to medical record
Before intervention[1]	Number of events					
Pre-postassessment group:						
Serious or very serious	18	12	6	1	5	67
In between	34	20	14	1	13	59
Minor or very minor	37	25	12	1	11	68
After intervention[2]						
Total:						
Serious or very serious	77	46	31	3	28	60
In between	101	71	30	4	26	70
Minor or very minor	98	68	30	3	27	69
Pre-postassessment group:						
Serious or very serious	31	21	10	1	9	68
In between	49	36	13	1	12	73
Minor or very minor	40	31	9	1	8	78
Postassessment-only group:						
Serious or very serious	46	25	21	2	19	54
In between	52	35	17	3	14	67
Minor or very minor	58	37	21	2	19	64

[1]7 recalled events that did not receive a seriousness rating (false positives elicited during free recall but later retracted) are not included.
[2]2 events for which a seriousness rating was not obtained (false positives added after free recall) are not included.

Next, we sought to ascertain whether this unexpected result could be attributed to a tendency to import serious events from outside the reference period, resulting in more false positives for events the subjects regard as serious. We checked the proportion of recalled events that were forward-telescoped false positives (events that occurred between 12 and 18 months ago, but which subjects reported to have occurred within the last year) for each seriousness group and found the rates to be 4, 4, and 3 percent, respectively, for the high, medium, and low event seriousness categories. Thus, failure to find more medical record matches for recalled events that subjects rated as serious does not appear to be an artifact of a greater incidence of forward-telescoped false positives for these events.

As mentioned above, the recall measure used elsewhere in our analyses (matches/medical record events) could not be used when evaluating differences related to subjects' assessments of event seriousness, because subjects gave no ratings for medical record events they did not recall. Accordingly, each recalled event and each event on the medical records were also evaluated as serious or minor by two experimenters, as was done in the pilot study (see appendix I). Using experimenter event seriousness classifications, recall accuracy was computed for serious and minor events as shown in table M. This analysis, like that of the subjects' seriousness ratings reported above, failed to reveal any advantage for serious events. In fact, there was a trend for more complete recall of minor events than serious events before the intervention (47 percent compared with 31 percent, $Z = 1.82$, $p < .07$). After the intervention, recall for serious and minor events did not vary significantly; the nonsignificant tendency for a greater proportion of

Table M. Recall before and after the intervention by subject group and experimenter's assessment of seriousness

Intervention status, subject group, and seriousness rating	Number of events on medical record (1)	Matches: Number (2)	Matches: Number per record event (3)	False negatives: Number (4)	False negatives: Number per record event (5)	False positives: Number (6)	False positives: Number per record event (7)	Net error rate[1] (8)	Net error[2] (9)
Before intervention									
Pre- and postassessment group:									
Minor events	85	40	0.47	45	0.32	27	0.53	0.85	0.21
Serious events	54	17	0.31	37	0.22	12	0.69	0.91	0.46
After intervention									
Total:									
Minor events	184	117	0.64	67	0.34	62	0.36	0.70	0.03
Serious events	126	68	0.54	58	0.25	31	0.43	0.71	0.21
Pre- and postassessment group:									
Minor events	85	55	0.65	30	0.27	23	0.35	0.62	0.08
Serious events	54	33	0.61	21	0.19	10	0.39	0.57	0.20
Post-assessment-only group:									
Minor events	99	62	0.63	37	0.39	39	0.37	0.77	–0.02
Serious events	72	35	0.49	37	0.29	21	0.51	0.81	0.22

[1]Gross error rate equals the sum of column 5 and column 7.
[2]Net error rate equals the difference between column 5 and column 7.

matches among minor events (mean for minor events, 64 percent, and for serious events, 54 percent; $Z = 1.77$, $p > .08$) is offset by a larger false positive rate. Thus we found no strong evidence that recall level is a function of event seriousness, as defined by either the experimenters or the subjects themselves.

Subject variables

The above analyses provide a detailed report of the recall data using the medical record event as the unit of analysis. Several additional analyses were done to compare the performance of different subjects and evaluate the effects of specific subject variables.

Subject variables thought to have a potential influence on recall in the free recall task were examined. Caution is required in interpreting these data because the sample sizes are quite small. The pre-postassessment group data before the intervention do suggest, however, that the proportion of events recalled was marginally different for men and women ($Z = 1.75, p < .10$), with women tending to report a greater proportion of matches (49 percent versus 31 percent).

Educational level significantly influenced the proportion of events remembered after free recall: Those with less than a college degree remembered 22 percent of the events on their medical records compared with 56 percent for those with a college degree or higher education ($Z = 3.41$, $p < .001$).

Another subject variable that we hypothesized would influence recall was having a chronic health condition. We assessed the efficacy of the intervention (using the pre-postassessment group) for those who did and those who did not have a chronic condition (as reported on the annual membership survey and confirmed by checking the medical record).

A 2×2 unweighted means repeated measures ANOVA (analysis of variance) was conducted, with assessment occasion (before or after the intervention) as the within-subjects factor and with health status (chronic condition, none) as the between-subjects factor. Table N shows mean recall performance for each group in the analysis. The intervention significantly improved recall for pre-postassessment group subjects (F (1, 16) = 7.80, $p < .05$). There was no main effect of having a chronic condition ($F < 1$), but the Chronic Condition X Assessment Occasion interaction was significant (F (1, 16) = 4.68, $p < .05$), with subjects who had chronic conditions showing larger gains as a result of the intervention than subjects who did not report having a chronic condition.

Table N. Mean proportion of medical record events recalled before and after the intervention, by subject's reported health condition

Condition[1]	Mean proportion recalled: Before intervention	Mean proportion recalled: After intervention	Difference
Chronic condition (n = 10)	37	64	27
No chronic condition (n = 8)	46	47	1
Total (n = 18)	41	56	15

[1]As reported by subjects and verified on medical records.

NOTE: Based on 18 subjects in the pre-postassessment group.

Dating

A second dependent measure investigated was the subject's ability to remember the dates of recalled health events. For each event, both the actual number of days that had elapsed since the event and, if the subject provided a date for the event, the difference between the actual date and the recalled date were computed. (For events for which the subjects could only remember a month, the 15th day of the month was used in computations. Similarly, the midpoint of the season was used in cases where the subjects recalled only the season in which the event had occurred.)

Table O shows the accuracy of dates recalled before the intervention for events of various recencies. The subjects were more likely to recall a date for more recent events, and recalled dates for these events were more accurate than those for older events. Table P shows the effect of the intervention on dating accuracy. Before the intervention, subjects in the pre-postassessment group recalled a date within 15 days of the actual date of occurrence for 19 percent of events on record. After the intervention, the proportion increased to 34 percent. The improvement in dating accuracy for the pre-postassessment group was significant (Cochran's Q (1, N = 139) = 14.29, $p < .01$), as was the difference between the proportion of events dated within 15 days of actual occurrence for the postassessment-only group (31 percent) contrasted to the before-intervention dating of the pre-postassessment group ($Z = 2.05, p < .05$).

Table O. Pre-postassessment group's dating accuracy before the intervention, by event recency

Time between event and interview	Number of events on medical record	Number of record events: Dated	Number of record events: Dated within 15 days	Number of record events: Forward-telescoped	Number of record events: Backward-telescoped	Percent of record events: Dated	Percent of record events: Dated within 15 days	Percent of record events: Forward-telescoped	Percent of record events: Backward-telescoped
All events	[1]139	53	27	15	11	38	19	11	8
0–3 months	30	21	11	7	3	70	37	23	10
4–6 months	42	20	8	4	8	48	19	10	19
7–9 months	29	7	5	2	0	24	17	7	0
10–12 months	38	5	3	2	0	13	8	5	0

[1]4 record events that subjects recalled but dated outside the reference period are included.

Table P. Dating accuracy before and after the intervention, by subject group and event type

Subject group and event type	Number of events on medical record	Number of medical record events: Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped	Percent of medical record events: Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped
Total:					Before intervention				
All events	309	...	...	...	...	...	...	...	...
Recurring events	153	...	...	...	...	...	...	...	...
Nonrecurring events	156	...	...	...	...	...	...	...	...
Pre-postassessment group:									
All events	[1]139	53	27	15	11	38	19	118	8
Recurring events	81	25	18	4	3	31	22	5	4
Nonrecurring events	58	28	9	11	8	48	16	19	14
Postassessment-only group:									
All events	[2]170	...	...	...	...	...	...	...	...
Recurring events	72	...	...	...	...	...	...	...	...
Nonrecurring events	98	...	...	...	...	...	...	...	...
Total:					After intervention				
All events	309	188	99	45	44	61	32	15	14
Recurring events	153	95	49	23	23	62	32	15	15
Nonrecurring events	156	93	50	22	21	60	32	14	13
Pre-postassessment group:									
All events	[1]139	92	47	24	21	66	34	17	15
Recurring events	81	56	29	14	13	69	36	17	16
Nonrecurring events	58	36	18	10	8	62	31	17	14
Postassessment-only group:									
All events	[2]170	96	52	21	23	56	31	12	14
Recurring events	72	39	20	9	10	54	28	13	14
Nonrecurring events	98	57	32	12	13	58	33	12	13

[1]4 record events that subjects recalled but dated outside the reference period are included.
[2]1 event was dropped because the correct date was unknown.

Table Q. Dating accuracy after the intervention, by event type

Event type	Number of events on medical record	Number of medical record events: Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped	Percent of medical record events: Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped
All events	[1]309	188	99	45	44	61	32	15	14
Total recurring events	153	95	49	23	23	62	32	15	15
Total nonrecurring events	156	93	50	22	21	60	32	14	13
Serious events	126	70	36	23	11	56	29	18	9
Recurring serious events	69	37	19	11	7	54	28	16	10
Nonrecurring serious events	57	33	17	12	4	58	30	21	7
Minor events	183	118	63	22	33	64	34	12	18
Recurring minor events	84	58	30	12	16	69	36	14	19
Nonrecurring minor events	99	60	33	10	17	61	33	10	17

[1]4 record events that subjects recalled but dated outside the reference period are included. 1 event was dropped because the correct date was unknown.

One issue of particular interest was the relationship between event seriousness and the subject's ability to date the event. Table Q shows the accuracy of remembered dates for the various types of health events. The data indicate that serious events were no more accurately dated than events regarded as minor (56 percent versus 64 percent, $Z = 1.58, p < .12$); in fact the difference was in the opposite direction. Next, we computed the number of forward- and backward-telescoped events. Forward-telescoped events were defined as events brought forward in time by 15 days or more. Backward-telescoped events were those for which the remembered date was 15 days or more before the actual date. There was no relationship between type of telescoping and the recurring-nonrecurring nature of the event: Forward- and backward-telescoping were equally common for both recurring and nonrecurring events. There was a relationship between forward-telescoping and event seriousness, however. Forward-telescoping was more common than backward-telescoping for serious events, whereas minor events exhibited more backward-telescoping. This provided weak support for the accessibility hypothesis of Brown, Rips, and

Shevell (1985), as we found marginally greater forward-telescoping for serious events (33 percent) than for minor ones (19 percent) ($Z = 1.77, p < .10$).

This finding contrasts somewhat with our pilot study results in which events categorized as serious by the experimenters were dated more accurately than minor ones. In the present study, the larger number of medical events in the lives of our subjects apparently made even relatively serious health-related events hard to date. (In the pilot study, 41 percent of all serious events on record were dated within 15 days of their actual occurrence compared with 28 percent of serious events before the intervention in the present study.) In addition, subjects in the present study had not experienced births or very serious operations, as had some subjects participating in the pilot study.

Health self-assessment

Subjects provided overall ratings of their health twice—once several days before the interview when asked to do so by the health plan staff member who set up the interview, and once at the conclusion of the interview. For half of the subjects, the five response options were arranged from excellent to poor; for the other half of the subjects, options were presented in reverse order, ranging from poor to excellent. Each subject was asked the question in the same order on both occasions. Regardless of question order, responses were coded as follows—excellent = 5; very good = 4; good = 3; fair = 2; poor = 1.

There were no significant differences between health assessments made in response to the plan staff member's question on the telephone and those made after describing all health plan visits during the interview. The mean rating for the first assessment was 3.8, and that for the second, 3.6 ($t(36) = 1.64, p < .20$). The correlation between the two sets of ratings was .70 ($p < .01$).

An effect was found for response option order. Those for whom responses were listed from most positive to most negative gave more positive self-assessments on the telephone (4.28 versus 3.26, $t(35) = 3.95, p < .001$), and also after the interview (3.89 versus 3.24, $t(38) = 2.63, p < .05$) than did those who were asked the question with options ranging from poor to excellent. As in the pilot study, in which all subjects received the assessment question with options ordered from excellent to poor, none of the subjects given the question in this form rated his or her health as anything less than "good," even after describing numerous medical visits during the interview. In contrast, 4 of 21 subjects (19 percent) who received the assessment question with options ordered from poor to excellent described their overall health as "fair." None of the subjects with the poor-to-excellent order assessed his or her health as "excellent."

No effect of the subjects' sex was evident in the self-assessment of their health. For the 20 female subjects, the mean assessment was 3.6; for the 20 males, the mean assessment was 3.5 ($t(38) = 0.36, p > .20$). The result was the same when looking only at the subjects who received the excellent-to-poor response order (mean for females, 3.8; mean for males, 4.0; $t(17) = 0.40, p > .20$), or only at those who received the poor-to-excellent order (mean for females, 3.3; mean for males, 3.2; $t(19) = 0.48, p > .10$).

To assess the impact of actual visit frequency upon health self-assessment, the sample was split roughly at the median, taking those with six or fewer plan visits as the low-frequency group and those with seven or more visits as the high-frequency group. With the excellent-to-poor response order, actual visit frequency showed no relationship with health assessment on either occasion. At the conclusion of the interview, the mean assessment for the low-frequency group was 3.88 and for the high-frequency group, 3.90. This finding agrees with the pilot study, in which all subjects received the self-assessment question in this order. However, with the poor-to-excellent order, subjects in the low-frequency category had a higher mean health assessment than those with more plan visits—3.70 versus 3.00 ($t(19) = 2.19, p < .05$). The same pattern of results was found for the telephone assessments before the interview. This suggests that the poor-to-excellent order may elicit more realistic self-assessments.

To analyze the effect of health assessment upon recall, we combined the "excellent" with the "very good" respondents and the "good" with the "fair" respondents. In terms of the proportion of medical record visits recalled, the resulting two self-assessment groups were very similar: Mean recall rate was 57 percent for the higher self-assessment group and 59 percent for the lower group ($t(38) = 0.30, p > .20$). In particular, we had expected those with high self-assessments to exhibit more false negatives, but this prediction was not supported. The mean number of false negatives for subjects in the higher self-assessment group was 2.76 compared with 3.53 for those in the lower group ($t(38) = 1.00, p > .20$).

Summary

Recall before intervention

- The subjects recalled 41 percent of the health plan visits on their medical records. Initially, the subjects recalled a lower proportion of the recurring events on their medical records—32 percent—than of the nonrecurring events—53 percent.
- The subjects were twice as likely to fail to mention a visit on their medical record (false negatives) as they were to describe a visit that was not on the medical record (false positives).
- Recall was more complete for those with college degrees and marginally so for women.

Recall after intervention

- The decomposition–time line intervention produced significant improvement in the percent of medical record events recalled: 63 percent for the prepostassessment group and 57 percent for the post-assessment-only group.

- Recall improvement was greater for those subjects with chronic conditions.
- The intervention reduced the number of false negatives without increasing the number of false positives.
- The intervention was most effective for recurring events.
- Events that the subjects rated as "serious" were no more likely than those they rated as "minor" to be matched to a medical record entry.

Dating accuracy

- The intervention increased the percent of events for which the subjects provided a date. It also led to an improvement in the percent of events dated accurately (within 15 days of actual date).
- Forward-telescoping (recalling an event as more recent than it actually was) was more common than backward-telescoping (recalling an event as less recent than it actually was) for serious health events, whereas backward-telescoping was more common for minor events.
- There was no relationship between type or frequency of telescoping error and the recurring–nonrecurring nature of the event.

Health self-assessment

- Self-assessments of overall health made on two occasions correlated .70.
- Response option order affected health assessments: Subjects for whom options were ordered from excellent to poor rated their health more positively than did those for whom options were ordered from poor to excellent.
- Actual visit frequency had a significant relationship to self-assessment only for those who received the poor-to-excellent response option order.
- Health assessment groups did not differ in proportion of medical record visits recalled.

Discussion

Our study's findings have implications of both practical and theoretical significance. Neisser (1986), Linton (1975), Brewer (1986), and Reiser, Black, and Kalamarides (1986) have all written about generic memories arising out of repeated exposure to the same event and theorized that, with time, individual incidents are harder to remember than the generic memory of events. There has been little empirical evidence to support this theoretical position, however. Our access to medical records in this investigation allowed us to document this phenomenon: The initial probability of recalling an incident from a group of recurring events was .32 compared with a probability of .53 for other incidents. Moreover, subjects' self-reports corroborated the generic memory interpretation: For 15 of the 40 event groups in subject records, the subject referred to the group as a generic whole and stated that there had been more than one such event but the individual events could not be recalled.

This phenomenon has several implications for survey research. First, such generic memories or specialized event scripts have the surface features of reports of single incidents. They may lead to underreporting because the respondent provides a generic memory but the interviewer construes it as a single incident. Alternatively, underreporting may occur if subjects attempt to calculate event frequencies by recalling individual incidents and summing them and they have greater difficulty accessing events related to generic memories. We did not obtain separate frequency estimates for recurring-event groups, but to the extent that subjects rely on the number of events they can remember in making frequency estimates, our data would suggest that the poorer recall for individual incidents of a recurring nature will lead to underestimates of event frequency.

The number of individual events the subject can access may serve as an anchor in estimating event frequency, in the manner described by Tversky and Kahneman (1974). In their classic series of studies, they found that when making estimations under conditions of uncertainty, subjects will start with some initial value and then make adjustments to arrive at a final answer, but the adjustments are usually insufficient. Thus, if people begin with different anchors, their final answers are likely to be different. Tversky and Kahneman found large differences in subjects' estimates for the product of $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ and those for $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$. The median estimate for the first sequence was 512, whereas that for the second was 2,250: Apparently people multiply the first few numbers in the sequence and then estimate the effects of succeeding computations. Subjects estimating the frequency of recurring events may go through this same process, starting with the number of individual incidents they can recall as an anchor. If the number of individual incidents they can access is typically too low, we would expect final estimates to be low also.

In addition to providing evidence for the lower accessibility of recurring events in autobiographical memory, our study demonstrated the efficacy of a cognitively based set of techniques for enhancing recall of such events. Memory for recurring events rose from a level of .32 before the intervention to .67 afterward for the pre-postassessment group. The improvement was even more dramatic—from 16 percent to 78 percent—for recurring events that were the target of decomposition as well as of the time line.

The success of our intervention demonstrates the utility of the "multiple pathways" approach to obtaining access to individual incidents in autobiographical memory, even in cases where those individual incidents have led to the construction of a generic memory. The fact that false positives did not go up significantly after the intervention suggests that the subjects were recalling or reconstructing real incidents, rather than simply responding to perceived experimenter demand by reporting more incidents indiscriminately. Thus, for practical purposes, the multiple pathways technique is useful in improving the quality of autobiographical memory reports.

From a theoretical perspective, however, we would like to know whether the individual events remain in memory, in addition to the stronger script for the generic version of the event, as Reiser, Black, and Kalamarides (1986) propose, or whether the individual memories actually become fused to become the generic memory and are no longer individually accessible, as Linton (1986) suggests. Although our intervention was based on the former theoretical supposition, and thus we sought to give the subjects cues to the distinguishing features of individual traces presumed to still reside in memory, it is also possible that the intervention works because it helps subjects reconstruct individual incidents from other information available to them.

The intervention was particularly successful in terms of improving recall accuracy for those subjects with a chronic condition. We selected half of our subjects from those

health plan members who reported having a chronic condition on the annual membership survey. These people are likely to have experienced multiple events of the same type during the past year, and initially they tended to recall an undifferentiated group of events in response to NHIS-type questions. Our intervention was designed with the goal of helping the subjects first to retrieve individual episodes of a given type, and then to fit those events, along with other events, into a chronological context: The time line was set up so that personal events could serve as landmarks for reconstructing a chronology. In addition, such events can often serve as cues triggering the memory of additional items. The decomposition and time line did in fact produce recall levels for these individuals that were as good as those for subjects without a chronic condition. The procedure used in this laboratory study appears successful enough to warrant consideration as a possible technique for inclusion in national surveys.

We acknowledge that the procedure is not without cost. It certainly would add to interview length (an average of 15 minutes was required to set up the personal time line into which health events could be inserted), and the extent to which it improves the validity of responses in an operational setting would have to be assessed through more applied studies and weighed against increased costs. Interview pacing has been found to have an impact on the effectiveness of a laboratory-based technique for enhancing recall accuracy (NCHS, 1989), and we would speculate that a relatively slow, deliberate pace on the part of the interviewer is vital for the intervention developed here.

Our findings also have implications concerning the accessibility hypothesis of Brown, Rips, and Shevell (1985). According to that hypothesis, people's perception of the recency of an event is influenced by how accessible the event is in memory. Thus, an event that had more emotional impact at the time of occurrence or that has been read about, discussed, or rehearsed more will seem more recent than an event that occurred at the same time but is less accessible in memory. Brown, Rips, and Shevell (1985) demonstrated this phenomenon for public events, such as assassination attempts on President Reagan and the Pope. Several theorists have proposed a similar mechanism in autobiographical memory (for example, Linton, 1986) and our results offer some support for this position. Serious events, for example, were marginally more likely to be brought forward in time. However, the effect is small in magnitude and not robust; in the pilot study serious events were more accurately dated than minor ones, and less susceptible to forward-telescoping, suggesting that dating accuracy is affected by something more than accessibility. Accessibility may affect recency estimates only when the subject has neither a stored date tag for the event nor sufficient information from which to make a strong inference about the date.

Although tangential to the central issues investigated in this study, our finding regarding the strong effect of response option order on the subjects' assessment of their overall health warrants discussion. Subjects for whom options were ordered from excellent to poor never rated their overall health as anything less than "good," the midpoint on the scale. Subjects who received options ordered from poor to excellent, on the other hand, never reported their overall health as "excellent" and often selected the "fair" option.

From a theoretical perspective, these results can be interpreted in terms of Tversky and Kahneman's (1974) model of anchoring and adjustment, described above. Our findings would be predicted if one assumes that respondents take the first option on a response scale as the anchor in estimating their response to the health self-assessment question. Danchik and Drury (1985) reviewed the factors that can affect responses to the health self-assessment question, such as number of response options or question placement. We believe that response option order should certainly be added to the list of factors that survey researchers consider. (Survey researchers have long known that response option order can affect responses, but option order is generally not counterbalanced across subjects for reasons of economy.) As long as the relationship between the item scale and other variables of interest is unchanged, survey analysts may not really care whether the average subject reports overall health in the "good" or the "very good" range. However, for items on which a ceiling or floor effect is likely (such as the health assessment question in these health-conscious times), the survey designer can manipulate the item's anchor in a way that will tend to minimize the problem. In addition, our finding that the relationship between health self-assessment and the actual health plan utilization rate was affected by response-option order suggests a need for a better understanding of how individuals are formulating their responses to the self-assessment question.

Finally, from the survey research perspective, the importance of our study is that the intervention produced large increases in the accuracy of recall of health care visits—an important statistic for measuring health care needs in the population and for planning for future needs. In light of differences between this laboratory study and in-home health surveys—for example, differences in respondent motivation and educational levels—these findings need replication under field conditions before the intervention is used in surveys. In addition, the intervention would need some modification to reduce administration time. Nevertheless, the findings strongly suggest that there is an exciting potential for adopting similar procedures in health surveys. The accuracy of a wide variety of survey-based statistics, not just those pertaining to health care visits, may be enhanced by the adoption of this and other cognitive psychological approaches in survey development.

References

Anderson, R. C. 1984. Some reflections on the acquisition of knowledge. Educational Researcher 13: 5–10.

Anderson, R. C., R. J. Spiro, and M. C. Anderson. 1978. Schemata as scaffolding for the representation of information in connected discourse. American Educational Research Journal 15: 433–440.

Baddeley, A. D. 1979. The limitations of human memory. In L. Moss and H. Goldstein, eds., The Recall Method in Social Surveys, pp. 13–30. London: University of London Institute of Education.

Barclay, C. R. 1986. Schematization of autobiographical memory. In D. C. Rubin, ed., Autobiographical Memory, pp. 82–99. Cambridge: Cambridge University Press.

Bower, G. H., J. B. Black, and T. J. Turner. 1979. Scripts in memory for text. Cognitive Psychology 11: 177–220.

Brewer, F. 1986. What is autobiographical memory? In D. C. Rubin, ed., Autobiographical Memory, pp. 25–49. Cambridge: Cambridge University Press.

Brown, N. R., L. J. Rips, and S. K. Shevell. 1985. Subjective dates of natural events in very-long-term memory. Cognitive Psychology 17: 139–177.

Brown, N. R., S. K. Shevell, and L. J. Rips 1986. Public memories and their personal context. In D. C. Rubin, ed., Autobiographical Memory, pp. 137–158. Cambridge: Cambridge University Press.

Danchik, K. M., and T. F. Drury 1985 Evaluating the effects of survey design and administration on the measurement of subjective phenomena, the case of self-assessed health status. Paper presented at the Annual Meeting of the American Statistical Association, Survey Research Section. August 5–8. Las Vegas, Nev.

Davies, G. 1986. Context effects in episodic memory, a review. Cahiers de Psychologie Cognitive 6: 157–174.

Geiselman, R. E., R. P. Fisher, D. P. MacKinnon, and H. L. Holland. 1985. Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. Journal of Applied Psychology 70: 401–412.

Jabine, T. B. 1985. Reporting Chronic Conditions in the National Health Interview Survey: A Review of Findings From Evaluation Studies and Methodological Tests. Unpublished report. U.S. Department of Health and Human Services. Rockville, Md.

Linton, M. 1975. Memory for real world events. In D. A. Norman and D. E. Rumelhart, eds., Explorations in Cognition, pp. 376–404. San Francisco: Freeman.

Linton, M. 1986. Ways of searching and the contexts of memory. In D. C. Rubin, ed., Autobiographical Memory, pp. 50–67. Cambridge: Cambridge University Press.

Loftus, E. F., and W. Marburger. 1983. Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. Memory and Cognition 11: 114–120.

National Center for Health Statistics, W. G. Madow. 1967. Interview data on chronic conditions compared with information derived from medical records. Vital and Health Statistics. Series 2, No. 23. PHS Pub. No. 1000. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, W. G. Madow. 1973. Net differences in interview data on chronic conditions and information derived from medical records. Vital and Health Statistics. Series 2, No. 57. DHEW Pub. No. (HRA) 76–1331. Health Resources Administration. Washington: U.S. Government Printing Office.

National Center for Health Statistics, B. Bloom. 1982. Current estimates from the Health Interview Survey, United States, 1981. Vital and Health Statistics, Series 10, No. 141, p. 49. DHHS Pub. No. (PHS) 82–1569. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, J. Lessler, H. C. Mitzel, W. Salter, and R. Tourangeu. 1989. Questionnaire Design in the Cognitive Research Laboratory: Results of an experimental prototype. Vital and Health Statistics. Series 6, No. 1. DHHS Pub. No. 89–1076. Public Health Service. Washington: U.S. Government Printing Office.

Neisser, U. 1986. Nested structure in autobiographical memory. In D.C. Rubin, ed., Autobiographical Memory, pp. 71–81. Cambridge: Cambridge University Press.

Reiser, B. J., J. B. Black, and P. Kalamarides. 1986. Strategic memory search processes. In D. C. Rubin, ed., Autobiographical Memory, pp. 100–121. Cambridge: Cambridge University Press.

Roediger, H. L., III, and D. G. Payne. 1982. Hyperamnesia: The role of repeated testing. Journal of Experimental Psychology: Learning, Memory, and Cognition 8: 66–72.

Schank, R. C., and R. P. Abelson. 1977. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Erlbaum.

Smith, S. M. 1979. Remembering in and out of context. Journal of Experimental Psychology: Human Learning and Memory 5: 342–361.

Sudman, S., and N. M. Bradburn. (1973). Effects of time and memory factors on responses in surveys. Journal of the American Statistical Association 68: 805–815.

Tulving, E. 1983. Elements of Episodic Memory. Oxford: Oxford University Press.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: heuristics and biases. Science 185: 1124–1131.

Appendixes

Contents

I. Pilot study 24
Recall performance 25
Recall organization 25
Dating accuracy 25
Self-assessment of overall health 26
Summary 26
II. Sample selection and characteristics 28
Selection of potential subjects 28
Recruitment of subjects 28
Response rates 29
III. Experimenter protocol 31
IV. Record matching and scoring 36
Contents and quality of medical records 36
Record matching procedures 36

List of appendix figures

I. Sampling and recruitment of George Washington University Health Plan subscribers 30
II. Time line 34

Appendix I
Pilot study

Subjects were 27 volunteers from the George Washington University Health Plan. Potential subjects were drawn from plan members between the ages of 18 and 65 years who had responded to the 1984 annual membership survey and who reported having had at least three primary care visits between September 1, 1984, and August 31, 1985. The resulting group of potential subjects was so stratified that half the potential participants were drawn from the lower half of the visit frequency distribution (3–4 visits) and half from the upper half (5–10 visits). The sample was also stratified on education: One-third of the subjects from each visit frequency group were individuals who had reported their educational level as high school graduation or less; two-thirds had reported some college education or more. Letters inviting participation and telephone contact procedures were essentially the same as those used in the major study, as reported in appendix II. Of 39 people contacted by telephone, 27 (69 percent) agreed to participate.

The sample obtained through this selection and solicitation process consisted of 9 men and 18 women, ranging from 24 to 62 years in age. The number of verifiable health plan visits on the medical record ranged from 0 to 14 with a mean of 4.30. (Unfortunately, the reference periods for the annual membership survey and for our interview were not identical, and three subjects who had reported three primary care visits on the membership survey were later found to have no verifiable office visits during the period covered by our questions.)

In our pilot study, we wanted to explore the strategies people use in recalling and dating health events of the kinds that appear in the National Health Interview Survey (NHIS). Our primary method for developing insights into subject strategies was to have the subjects think aloud as they tried to recall all such events for the past year and to remember their date of occurrence. In addition, for 15 subjects, after they had recalled as many events as they could, we asked them to sort the events into two or more groups of incidents that go together in their own minds. They were then asked to describe the justification or reasoning underlying their categorization.

The pilot study explored the possibility that memory processes would differ, depending on whether the to-be-remembered event was a recurring event or a unique one, and whether the event was serious or minor in nature. Both of these classifications were made by the researchers. An event was labeled *recurring* if it involved a condition that had triggered at least three medical events during the past year. Events caused by other conditions were classified as *nonrecurring*.

A second type of classification concerned the seriousness of the health incident. An event was classified as *serious* if it involved a problem that would have a high probability of resulting in complications such as a major infection, debility, or death without the intervention or assistance of a medical professional. All other events were classified as *minor*. The distribution of subjects' medical record events by recurrence and seriousness is shown in the first column of table I. Overall, agreement among rates for classifications of seriousness was 88 percent. Disagreements were resolved through further discussion of individual events.

In addition, the pilot study explored the possibility that the subject's schema concerning his or her own health might affect the way in which memory was searched in response to NHIS-type questions. Subjects were asked to assess their health on a five-point scale, ranging from

Table I. Pilot study recall by event type

Event type	Number of events on medical record	Number: Matches	Number: False positives	Number: False negatives	Percent: Matches	Percent: False positives	Percent: False negatives
All events	[1]114	53	49	61	46	43	54
Total recurring events	54	20	18	34	37	33	63
Minor recurring events	26	7	8	19	27	31	73
Serious recurring events	28	13	10	15	46	36	54
Total nonrecurring events	60	33	31	27	55	52	45
Minor nonrecurring events	44	23	27	21	52	61	48
Serious nonrecurring events	16	10	4	6	63	25	38

[1]Does not include 1 event that could not be classified.

excellent to poor, either before or after recalling as many health plan visits as they could for the 12-month period preceding the interview.

Recall performance

As shown in table I, 46 percent of events on the medical records were recalled. Nonrecurring events were more likely to be recalled than recurring events (55 versus 37 percent). However, there was an interaction between the type of condition involved and the seriousness of the incident. Although minor health events that were nonrecurring were remembered almost as well as serious nonrecurring events (52 versus 63 percent), minor recurring events were least likely to be recalled (27 percent compared with 46 percent for serious recurring events).

Additional evidence suggesting that recurring medical events are recalled differently from nonrecurring events is found in the error data. For recurring events, there were more false negatives than false positives (63 compared with 33 percent). For nonrecurring events, the rate of false negatives and that of false positives were similar (45 and 52 percent, respectively).

Recall organization

The organization of autobiographical memory was explored both through a review of the think-aloud protocols and through use of the event-sorting data.

One of the most striking aspects of the subject protocols was the fact that many subjects (44 percent) had some class of events, such as dermatologist visits or days home sick from work, for which they asserted that there were multiple instances over the last 12 months, but they couldn't say how many or describe them individually. Our subsequent reading of Neisser's (1986) description of the "nested" character of autobiographical memory provided a nice conceptual framework for this phenomenon—Neisser's concept of "molar memories" appears to be an appropriate description of many of the responses elicited by NHIS-type questions.

Use of verbal protocols to identify the subjects' retrieval strategies was hampered by the frequency of NHIS-type experimenter prompts (for example, "Were there any other doctors, such as dermatologists, or eye doctors?") that tended to direct the recall process. In those instances when stretches of uninterrupted subject recall were obtained, however, we did an informal analysis that identified the following three types of retrieval organizations. Some subjects recalled events *chronologically,* ordering them by time of occurrence. Two subjects started with the first event in the reference period; one started with the most recent event and worked backward. In several additional cases, subjects appeared to embark on a temporal ordering of their recall, only to get side-tracked and start focusing on one particular type of event. Organizing events by type or *condition* was a common recall strategy, and one that would be compatible with the current NHIS format. Perhaps the most interesting retrieval strategy observed was one we call the *condition narrative;* here, multiple events related to a particular illness or injury are recalled in a storylike structure. There is temporal order, but only those events related to the narrative theme are included.

Dating accuracy

For the 115 events on the medical records, dating accuracy was computed by (1) event recency and (2) event type. As table II shows, only 25 percent of events on the medical records were recalled and dated within 15 days of actual occurrence. Those events that were accurately dated tended to be of recent vintage: 43 percent of events that had occurred within the last 3 months were dated within 15 days of actual occurrence, compared with 26 percent of events 4–6 months old, 13 percent of events 7–9 months old, and 24 percent of events 10–12 months old.

The data in table III suggest that event seriousness affects dating accuracy also: 41 percent of serious record events were dated within 15 days of the actual date compared with just 16 percent of minor events. The poorest dating by far was observed for recurring minor events (4 percent dated within 15 days of actual occurrence).

The think-aloud protocols, combined with the small number of events for which precise dates were remembered, suggest that people do not store date tags with most autobiographical events. Rather, they appear to use a system of landmark events, for which they have at least an approximate date, and to estimate dates for other events based on relationships with the landmark event. The subjects used holidays, weddings, parties, job changes, and major health events as dating landmarks. Of the 115 medical events on record for the pilot subjects, there was perfect recall of dates for 8 events. Three of these were events that had occurred within 2 weeks of the interview; these were of less interest than older events for which exact dates were

Table II. Pilot study dating accuracy by event recency

Time between event and interview	Number of events on medical record	Number of medical record events				Percent of medical record events			
		Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped	Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped
All events	115	47	29	10	8	41	25	9	7
0–3 months	23	16	10	3	3	70	43	13	13
4–6 months	23	12	6	2	4	52	26	9	17
7–9 months	32	8	4	3	1	25	13	9	3
10–12 months	37	11	9	2	–	30	24	5	–

Table III. Pilot study dating accuracy by event type

Event type	Number of events on medical record	Number of medical record events				Percent of medical record events			
		Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped	Dated	Dated within 15 days	Forward-telescoped	Backward-telescoped
All events	[1]114	47	29	10	8	41	25	9	7
Total recurring events	54	17	13	3	1	31	24	6	2
Total nonrecurring events	60	30	16	7	7	50	27	12	12
Serious events	44	22	18	2	2	50	41	5	5
Recurring serious events	28	13	12	1	–	46	43	4	–
Nonrecurring serious events	16	9	6	1	2	56	38	6	13
Minor events	70	25	11	8	6	36	16	11	9
Recurring minor events	26	4	1	2	1	15	4	8	4
Nonrecurring minor events	44	21	10	6	5	48	23	14	11

[1]Does not include 1 record event that could not be classified.

remembered. In this latter category were three emergency room visits, one birth, and one annual physical examination. All but the last could be described as serious and nonrecurring. Inspection of the protocol in which the date for the annual physical examination was recalled showed that the subject reported that he gets an annual physical on the same date every year.

In addition to these health events for which dates were perfectly remembered, there were several cases in which a health event that had occurred before the reference period served as a dating landmark and retrieval cue for health events that were within the reference period ("... on October 17 of '84, which is outside of this period, I had a triple [bypass] operation and stayed in the hospital 7 days ... and thereafter ... I came back for follow-up checkups ...").

Self-assessment of overall health

When asked to rate their overall health on a five-point scale ranging from "excellent" to "poor," none of the 27 pilot study subjects described his or her health as anything less than "good." Health self-assessments were not only strongly positive overall, they were also surprisingly independent of the number of health plan visits on record. The mean number of office visits was 4.29 for the "excellent" group, 4.92 for "very good," and 3.14 for the "good" group. Moreover, several subjects with quite serious health conditions evaluated their health as "excellent" or "very good." Such findings raise questions concerning what is being measured by the self-assessment item. To try to shed some light on how the subjects were interpreting this question, we asked them what kind of health they had considered in their answers (physical, mental, or both) and what they had used for a reference (for example, all people or their own prior health). Roughly half of the subjects said they considered both physical and mental health, whereas the other half said they thought only about physical health. About half of the subjects said they were using their own prior health as a reference in making their assessment; about half said that they compared themselves to all other people. About a third of the subjects defined a more specific reference group in terms of people their own age, or people their own age and sex (multiple responses were permitted).

Women tended to give more positive self-assessments than men, and more highly educated people tended to be more positive than the less educated, but neither difference was statistically significant.

We explored the relationship between health self-assessment and recall of health events for the past year. The mean proportion of record events recalled was .45 for those who regarded their health as "excellent," .55 for the "very good" group, and .42 for those who described their health as "good," showing no monotonic relationship between self-assessment and recall. We did find a significant difference between health assessment groups in the likelihood of having a molar memory—71 percent of those describing their health as "excellent" had one or more such memories, compared with 38 percent of those regarding their health as "good"($\chi^2 = 11.94, p < .01$).

Summary

Although limited by the small sample size, the pilot study produced several interesting trends. Recall for health events that had occurred during the past 12 months averaged 46 percent correct. Nonrecurring medical events were more likely to be recalled than recurring ones, and serious events were better recalled than minor ones. There appeared to be an interaction such that event repetition had a larger effect on memory for minor events (27 versus 52 percent) than for serious ones (46 versus 63 percent).

One of the most striking aspects of the recall protocols was the prevalence of what we termed *molar memories.* These were event clusters for which the subject described the type of event, reported that there had been multiple events of the type, but could not describe specific incidents. In all, 44 percent of the pilot study subjects reported one or more of these "molar" event groups. Often these reports were accompanied by remarks concerning the impossibility of recalling individual events because they were "all the same." Thus, the protocol findings suggest that subjects have difficulty recalling individual events that are instances of a category of recurring events. These often involved recurring conditions, although cases of molar memories for routine visits and acute conditions occurred as well.

Precise dates were recalled for only 8 of the 47 events matched to medical records for which subjects provided dates. With one exception, perfectly remembered dates were either very recent or serious nonrecurring events. Remembered dates for major health care events or for other personal incidents from the past year were used as landmarks in estimating dates for other events. Dating accuracy was better for serious events than for minor ones. Among minor events, dating was better for nonrecurring events than for recurring ones. Event repetition had a smaller effect on dating performance for serious health events.

When asked to rate their overall health, the pilot study subjects gave very positive self-assessments, using only the three most positive ratings on the five-point scale. Perhaps because of this restriction in range, there was no logical relationship between health assessment and frequency of health plan visits. Health self-assessment was not related to the proportion of medical-record events recalled. There was a significant difference in the number of "molar memories" reported, however, with these generic memories being more common among those who assessed their overall health as excellent.

Appendix II
Sample selection and characteristics

Selection of potential subjects

The subjects were recruited from the membership of the George Washington University Health Plan, which is a federally qualified health maintenance organization located in Washington, D.C. It is affiliated with the George Washington University and American Medical International. Its 23,000 members receive all primary medical care from its downtown Washington facility or its satellite facilities in suburban Maryland and Virginia.

The health plan maintains a single medical record for all patient encounters, laboratory, x ray, and consultant requests and reports. Hospital and emergency room notification and discharge summaries are filed in the medical record. Dates of admission as well as other pertinent data are also posted in the financial data base of the health plan.

Because a vital part of this study was the verification of recall for health events over the past year against health plan records, only individuals who had joined the health plan before January 1, 1985, were considered as potential subjects. To ensure that they would have a reasonable number of incidents to report, the target population was restricted further to those members who had had at least four primary care physician visits between August 1, 1985, and July 31, 1986 (roughly the 12 months before their interview). Following the respondent rule of the National Health Interview Survey (NHIS), only those 18 years of age or older were eligible and an age limit of 65 years was used. Most importantly, because we wanted to study memory for repeated health care incidents of the same type, we decided to select half our sample from those plan members with a recurring health problem.

Initially, a list of health plan members likely to meet our sampling criteria was culled from the responses to the plan's annual membership survey sent by the plan's Quality Assurance Program to a random sample of the plan membership each September. The survey included an item concerning the presence of an ongoing (chronic) health problem as well as a self-report on demographic information (age, sex, race, education) and the number of primary care visits during the past year. Potential subjects were those who reported four or more primary care visits and who met our other sampling restrictions (see below). Medical records were requested for these subjects and the actual number of total health plan visits during the period August 1, 1984, through July 31, 1985, determined by a plan staff member. Health plan members who had had four or more visits (primary or specialty) on record were retained as potential subjects. At this time, records for potential subjects who had reported having a chronic condition were checked to make sure that there was evidence of a condition always coded as chronic in NHIS (see table B) or one for which the individual had three or more health plan visits during the 12-month reference period. Because births appear to be uniquely memorable, subjects who had given birth during the past year were excluded from the sample.

To provide a balanced sample in terms of sex, utilization rate, and education level, potential subjects in each health condition stratum ("chronic" versus "nonchronic") were further stratified by sex (male, female), educational level (B.A. or more, less than B.A.), and total visit frequency (4–5 versus 6 or more). This sampling design guided the solicitation of plan members, from the eligible pool, for participation in this research. So far as possible, an equal number of subjects in each of the 16 groups defined by these variables was obtained. A concerted attempt was made to send letters to only those members who matched current needs for individual cells in the sampling design, to avoid creating expectations that would not be fulfilled.

Recruitment of subjects

The 81 recruitment letters were mailed in groups estimated to be of sufficient size to allow oversampling as required but not so great that the recruiter would not be able to contact and schedule interviews within a week or two of the respondent's receiving the letter.

The letter explained the purpose of the study, outlined the procedures to ensure the confidentiality of the information, requested the recipient's participation, specified a $20 compensation for time, and included the name and the telephone number of the health plan's Director of Quality Assurance if the recipient had any questions.

Approximately 7 days after mailing the recruitment letter, a health plan research assistant called each subject using the work or home telephone number listed in the registration file, supplemented as required with directory assistance.

Our standard locating procedure required that the health plan research assistant attempt to telephone each potential subject at least three times, at three different times of day, including morning, afternoon, and evening or

weekend. In practice, it was usual to make a dozen or more attempts and requests for forwarding numbers.

When the research assistant reached the subject, the assistant introduced herself and asked whether the recruitment letter had been received. If not, the assistant offered to mail another letter or to read the letter. If a new letter was requested, the assistant checked the address with the respondent and confirmed that she would call again in a week. If the subject requested that the letter be read, the assistant did so and continued with the scheduling protocol, asking if the subject would be willing to participate. For questions about the project's purpose or privacy, the assistant was instructed to repeat the pertinent paragraphs in the letter of recruitment.

The assistant then scheduled the time and place for the interview—at the subject's workplace or home, or at the offices of the health plan or Human Resources Research Organization (HumRRO). If the interview was scheduled for the workplace or home, traveling instructions were obtained. If the interview was to be at the HumRRO headquarters, the subject was given instructions and told that a map would be mailed.

The assistant then asked a general health self-assessment question: "How would you rate your overall health?" The order of response options (poor–fair–good–very good–excellent) was alternated across subjects.

After obtaining a telephone number for a telephone reminder and asking the subject to call the Quality Assurance Program if he or she needed to cancel or reschedule the interview, the assistant asked for any further questions and closed the call.

After scheduling an interview, the assistant called the HumRRO project headquarters with pertinent information and mailed a letter confirming each interview date, time, and location. A confirmation letter was sent to each scheduled subject noting the date, time, and location, and telephone number to call for changes. The letter also included an "assurance of confidentiality" statement.

Response rates

Figure I summarizes the selection, recruitment, and response rates. Sixteen percent (8 of 51) of those subjects who were contacted declined to participate in the study. Seven of those who declined were female and one was male. Of the 43 who were scheduled for an interview, only one (2 percent) canceled. Data from two additional subjects were dropped because of incompleteness: One subject broke off the interview before completion, and data from one interview had to be discarded because of experimenter error.

Table IV shows the characteristics of the 40 final subjects ("volunteers") compared with those of the 8 potential subjects who were contacted but declined to participate ("nonvolunteers"). This comparison is based on the self-report responses to the health plan membership survey.

Table IV. Characteristics of volunteers and nonvolunteers

Characteristic	*Volunteers (N = 40)*	*Nonvolunteers (N = 8)*
	Years	
Mean age	44.1	50.5
	Percent	
Race, black	40.0	*75.0
Sex, female	50.0	*87.5
Education, College—any	62.5	37.5
Chronic health problem	60.0	75.0
Health status—no activity limits, no medications	42.5	25.0

Note: *$p < .10$.

Table V. Characteristics of experiment subjects and other annual membership survey respondents

Characteristics	*Experiment subjects (N = 40)*	*Other annual membership survey respondents (N = 621)*
	Years	
Mean age	44.1	43.4
	Percent	
Race:		
Asian	–	1.2
Black	39.5	33.0
White	60.5	65.8
Sex:		
Female	50.0	56.8
Education:		
Less than college degree	37.5	33.7
College degree	25.0	25.1
Post-college graduate study	37.5	41.2
Chronic health problem	60.0	48.0
Health status:		
No activity limits, no medications	42.5	56.1
No limits, but on medications	47.5	32.7
Activity limits due to health	10.0	11.2

As shown in table IV, those who did not volunteer were more likely than volunteers to be older, black, and female, and were less likely to have attended college. However, none of the differences between volunteers and nonvolunteers was significant at the .05 level.

Table V shows a comparison of the 40 subjects in the final interview pool with the annual membership survey respondents, from which the 40 were drawn. The study sample matches the larger group of survey respondents rather closely, except for the deliberate oversampling of those who reported chronic conditions. It should be noted, however, that the survey respondents, like the health plan membership as a whole, are well above the national average in terms of educational level.

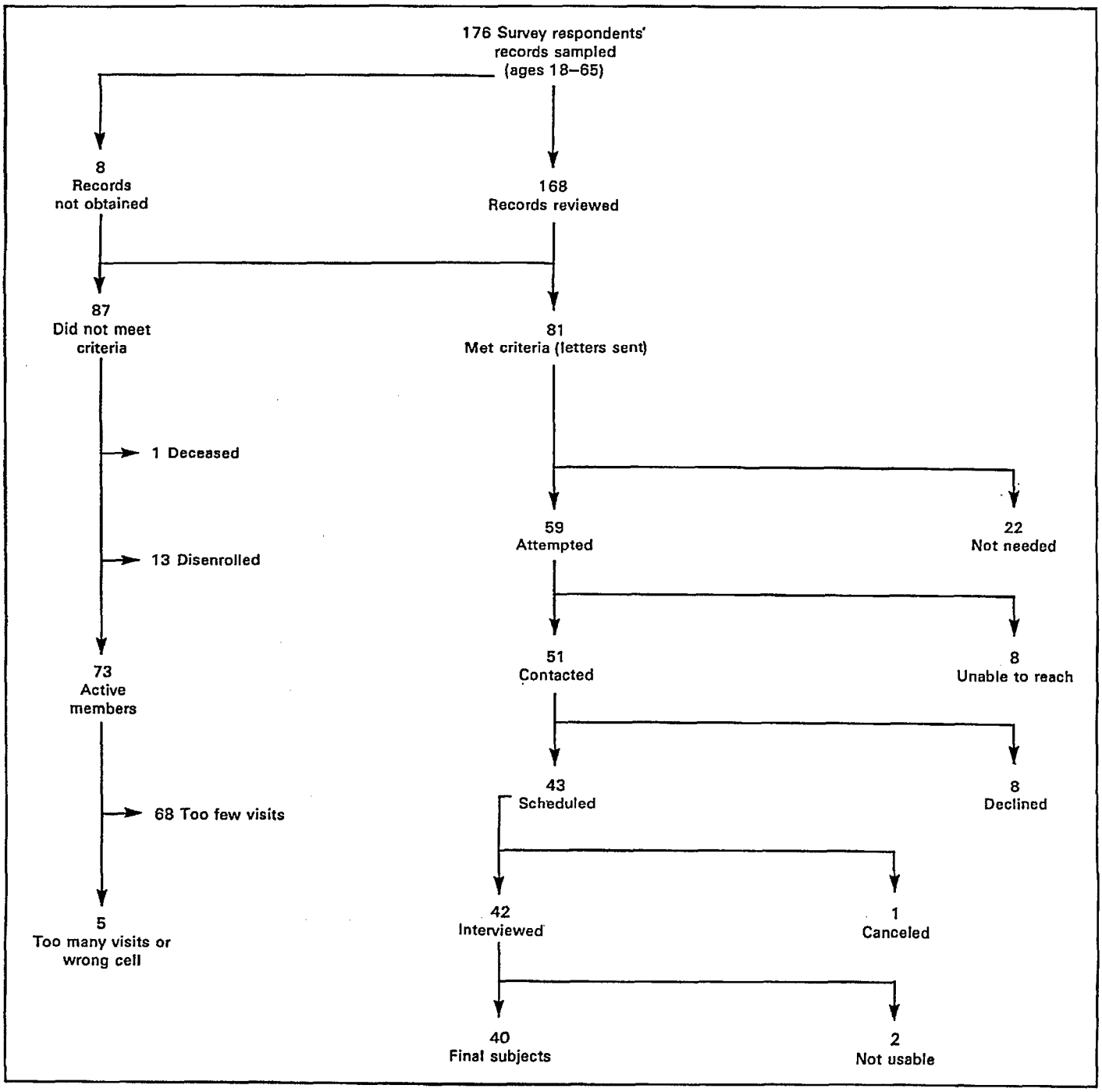

Figure I. Sampling and recruitment of George Washington University Health Plan subscribers

Appendix III
Experimenter Protocol

Brief Overview

[Explain project briefly.]

[Have respondent read Assurance of Confidentiality and sign the participation form.]

Mark Starting Time

[Write down starting time.]

[Start tape.]

Background Questions

During the last year, what was your major activity: working in a job outside of the home, going to school, or something else? **[Record primary activity.]**

[If job:] What sort of work do you do?

[If school:] What are you studying?

[If something else:] And what is that?

Estimate 1

Later on, I will ask you to describe your medical visits in the last year. Right now I'd just like you to give me a number. How many times, since (month) first, 1985 up until now, did you visit someone at the GW health plan about your health? Include medical doctors, nurse practitioners, physician assistants, hospital visits, and so on. **[DO NOT LET THEM START TO DESCRIBE; say you will want them to describe later.]**

Practice Concurrent Protocol

Now I will be asking you about specific health-related experiences you may have had in the past, about times that you may have been sick or injured, or gone to the hospital, or emergency room. As you remember those events I will ask you to think aloud, and describe the thoughts that lead you to remembering them.

Recalling medical visits is difficult in itself, and thinking aloud complicates the task. To help you get used to thinking aloud, I'll give you a few practice examples.

1. First, I want you to think aloud as you try to remember and describe when you last ate at a sit-down restaurant. Remember to tell me what's going through your mind as you answer the question.

2A. Now think aloud as you answer this question: How many times have you been to the grocery store in the past week?

[If long pause] What were you thinking before you started to speak?

2B. Describe the first time you were at the grocery store in the past week. Tell me everything that comes to mind.

3A. Okay, now think out loud as you answer this question: Have you watered a plant any time in the last 2 weeks?

3B. Describe aloud the most recent time. Again, tell me whatever comes to mind as you answer the question.

[USE BLUE PEN ONLY NOW.]

Free Recall

Now that you've had some practice thinking aloud, we're ready to go on. As you try to recall each event, remember to think aloud just as you did on the practice. Think carefully and try to remember every medical visit. Specifically, we're interested in the 12 months since (the first of today's month) a year ago, up to today.

As you know, your health plan includes many kinds of health professionals, including medical doctors, nurse practitioners, physician assistants, and so on. I want you to include all of these types of people as you respond. Do you have any questions before we begin? **[Pause for questions.]**

MAKING CARDS [Instruction to Interviewer]

[Take a stack of index cards for Free Recall.]

[As subject recalls events, write down a minimal number of words just to identify each recalled event or multiple-event.]

[Write down any factual information recalled, as well.]

[Make a separate card for each event recalled.]

[A health event would be any visit to a medical professional, or a visit to or stay in a hospital.]

[BE CAREFUL NOT TO FORGET FREE RECALL PROBES, BELOW.]

OMB#0937–0140
exp. 7/31/87

Free Recall

Now I want you to try to remember and describe out loud each of the times that you saw someone at your health plan about your health. Think aloud and describe each visit. (Start wherever you like.)

[For each incident, use prompts as follows:]

What brought that incident to mind?
Can you describe anything more about that visit?
How did you remember that?

[Prompt enough to get clear idea of respondent's strategy—what the respondent is remembering about the incident, what information is connected to what. DON'T PROMPT FOR FACTS; give nonspecific prompts.]

[For clumps: DON'T ask "Do you remember any particular visits for this problem?"]

Can you remember any more incidents?
How did you go about remembering these visits? Did you remember them in any particular order?

Estimate 2 (After Free Recall)

Now that you have told me about each visit, how many visits, then, would you say you've had in all, over the last 12 months? (I need your estimate *after* you've remembered the events.)

[If they give a range with less than four possibilities] If you had to choose, would you say _____ **[enumerate]**

[If they give a larger range] If you had to give a single number, how many visits would you say?

[Assign subject to experimental condition based on whether subject recalled multiple visits (1A or 1B alternately) or if no multiple visits recalled then assign subject to 2A or 2B alternately. From now on follow instructions for that group.]

Multiple visits

1A	1B
____ free recall	____ free recall
____ info cards	____ decomposition
____ decomposition	____ time line
____ time line	____ info cards
____ info cards	

No multiple visits

2A	2B
____ free recall	____ free recall
____ info cards	____ time line
____ time line	____ info cards
____ info cards	

[Group 1A or 2A]

Complete Information Cards After Free Recall

Now for each visit that you recalled, I need you to give me specific information.

You mentioned (visits/a visit that you went to the hospital) for (problem). What were the symptoms that led you to make the visit?

[Probes:]
What were the symptoms when you went in?
What was the reason for the visit? **[GET THE IMMEDIATE REASON OR SYMPTOM "ACCORDING TO THE SUBJECT."]**
[If appropriate] Was this connected to any other medical problem?

What did the problem turn out to be?
Was there a diagnosis?
Do you know what caused it?

Did they do any sorts of tests?
How about x rays?

What treatment did they suggest?
Did they suggest any medication?
What medicine?

Now, what was the date of this visit?
(Can you give me the month?)

This was at the GWU health plan?

Whom did you see? Was it a physician assistant ... medical doctor ... nurse practitioner ... medical student?

Do you remember what the name was?

Was this with a member of a primary care team ... specialty team ... emergency room ... hospital?

[NOW REPEAT THIS FOR NEXT EVENT CARD.]

[1A, 1B]

[USE GREEN PEN ONLY HERE.]

Decomposition Intervention

[For multiple-event cards:] All right, I noticed that this card really represents several visits **[REFER TO WHATEVER IS ON CARD]**. I'd like to know more about each one of them. Most people find such incidents difficult to remember, but I'd like you to try. Try to remember the last time that you made a visit for ___ . Can you describe anything about it?

[Do you need to Induce Recall or Enhance Recall?]

[To induce recall: subject-generated guided imagery] Now I want you to do some guided imagery. I want you to try to think about this visit, and imagine the place in the health plan building that you went to. I want you to picture the rooms and describe aloud everything you remember.
Now can you describe anything more about the visit itself?

[To enhance recall: context reinstatement] Do you remember anything else that happened that day?
Did you take off from work?

Do you remember how you got to the health plan offices?

Do you remember anything else you did that day?

Do you remember anything else that happened the day of the visits?

Did you have to wait long?

Was anyone else with you?

Was there anything different about that visit compared to the others?

[FOLLOW RESPONDENT'S CUES FOR ADDITIONAL PROMPTS.]

[When respondent is finished] When did you *first* make a visit for this problem?

[If within the last year, or 18 months] Now try to remember your first visit in the last year for . . . **[repeat cues]**

[When respondent is finished]

Now, can you remember anything about any of the other visits? **[CONTINUE AS LONG AS RESPONDENT CAN REMEMBER SOMETHING ABOUT INDIVIDUAL VISITS.]**

Estimate After Decomposition

Now that you have told me as much as you can about each visit for this condition, how many visits, then, would you say you've had in all, for the year?

[Do NOT let subject update info cards here—"We'll do that later."]

[1A, 1B, 2A, 2B]

[USE RED PEN ONLY HERE.]

Individual Time line Intervention [See figure II.]

Now I'd like you to work with this calendar. **[Give calendar and red pen to subject.]** It shows the 18 months leading up to now. **[Show on calendar.]** These are the 12 months we have been working with so far, and these are the 6 months before it. I'd like you to mark on the time line the dates for which you remember specific events. Circle the date and write down whatever you remember about that day underneath the month. These will be any personal landmarks, such as a birthday, starting a new job, a serious event that you would remember the date of, a wedding, a vacation, a holiday, or any other event for which you are pretty sure about the date. You can start anywhere you want, and try to put in as many landmarks as you can.

I want you to take as much time as you need for this, and remember as many dates as you can.

[Turn off recorder unless subject is talkative.]

[PROMPT FOR DIFFERENT TYPES OF LANDMARKS. Suggest obvious ones, for example, "Do you remember what you were doing on New Year's Day?" **TRY TO PROMPT FOR AT LEAST ONE LANDMARK IN EACH MONTH OF THE LAST YEAR.]**

[Turn on recorder.]

Insert Incidents in Time Line

Now first of all, I'd like you to think about these 6 months here **[first 6, before 12-month period]**, and try to remember if any of your medical visits happened in this time period.

[HAVE CARDS VISIBLE TO SUBJECT.]

Now I'd like you to think about (month) a year ago, looking at this time line with your landmarks on it, and try to remember if any of your health plan visits occurred in this time. Circle the date of the visit, and write it down under the month.

[Allow respondents to enter events outside the reference period or in other months as they choose.]

[1A, 1B, 2A, 2B]

Did you make any other visits to the health plan in this month (around ____ landmark), that you can remember? **[If a new visit, get card information, as before; and circle the new visit card number on the time line.]** Did it occur before or after (landmark)?

[Make a new card for any newly recalled incidents. Write down a few words that will identify that specific visit. Do this for each new visit.]

[Repeat with next month, even if all cards are done.]

[Allow respondents to enter events outside the reference period or in other months as they choose.]

[If there are any multiple-event cards left, then for each of these cards go through each month on the time line:] Do you think you made a visit for this problem in this month? Could you circle the date and write it down under the month?

[If after going through all months any event cards are left unrecorded in the time line:] Where on the time line should this visit be placed? **[Prompt for whether it was before or after certain landmarks, including landmarks outside the 1-year reference period.]** Do you want to guess the date?

[Review the medical events marked on the time line, to make sure they are clear.]

Estimate After Time Line

[Count up all visits in time line in the 1-year reference period.] All right; so here you have entered ____ visits during the last year. Then is that your best estimate of your visits in the last year?

[Group 1B or 2B]

[USE RED PEN ONLY HERE.]

Complete Information Cards After Time Line

Now for each visit that you recalled, I need you to give me specific information.

1985

JUNE
14 Flag Day
16 Father's Day

JULY
1 Canada Day (Canada)
4 Independence Day

SEPTEMBER
2 Labor Day
16 Rosh Hashanah
25 Yom Kippur

OCTOBER
12 Columbus Day
14 Columbus Day-Obsvd.
14 Thanksgiving Day (Canada)
24 United Nations Day
31 Halloween

NOVEMBER
5 Election Day
11 Veterans Day
28 Thanksgiving Day

DECEMBER
8 Hanukkah
25 Christmas

1986

JANUARY
1 New Year's Day
20 Martin Luther King

FEBRUARY
12 Lincoln's Birthday
12 Ash Wednesday
14 Valentine's Day
17 Washington's Birthday Obsvd.
22 Washington's Birthday

JUNE

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

JULY

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

AUGUST

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

SEPTEMBER

S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

OCTOBER

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

NOVEMBER

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

DECEMBER

S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

JANUARY

S	M	T	W	T	F	S
		1	2	3	4	
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

FEBRUARY

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	

1986

MARCH
17 St. Patrick's Day
23 Palm Sunday
28 Good Friday
30 Easter Sunday

APRIL
24 Passover

MAY
11 Mother's Day
17 Armed Forces Day
19 Victoria Day (Canada)
26 Memorial Day - Obsvd.
30 Memorial Day

JUNE
14 Flag Day
15 Father's Day

JULY
1 Canada Day (Canada)
4 Independence Day

SEPTEMBER
1 Labor Day

OCTOBER
4 Rosh Hashanah
12 Columbus Day
13 Columbus Day-Obsvd.
13 Yom Kippur
13 Thanksgiving Day (Canada)
24 United Nations Day
31 Halloween

NOVEMBER
4 Election Day
11 Veterans Day
27 Thanksgiving Day

MARCH

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

APRIL

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

MAY

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

JUNE

S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

JULY

S	M	T	W	T	F	S
	1	2	3	4	5	
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

AUGUST

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

SEPTEMBER

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

OCTOBER

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

NOVEMBER

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

Figure II. Time Line

You mentioned (visits/a visit/ that you went to the hospital) for (problem). What were the symptoms that led you to make the visit?

[Probes:]

What were the symptoms when you went in?
What was the reason for the visit? **[GET THE IMMEDIATE REASON OR SYMPTOM "ACCORDING TO THE SUBJECT."]**
[If appropriate] Was this connected to any other medical problem?

What did the problem turn out to be?
Was there a diagnosis?
Do you know what caused it?

Did they do any sorts of tests?
What were they?
How about x rays?

What treatment did they suggest?
Did they suggest any medication?
What medicine?

Now, what was the date of this visit?
(Can you give me the month?)

This was at the GWU health plan?

Whom did you see? Was it a physician assistant . . . medical doctor . . .nurse practitioner . . . medical student?

Do you remember what the name was?

Was this with a member of a primary care team . . . specialty team . . . emergency room . . . hospital?

[NOW REPEAT THIS FOR NEXT EVENT CARD.]

[Group 1A or 2A]

[USE RED PEN ONLY HERE.]

Update Information Cards After Time Line

[Give each event card to subject, one at a time.] Can I change or add any of the information on this event, or is it correct as it is? **[Check to make sure card date now corresponds to date on time line.]**

[1A, 1B, 2A, 2B]

Rate Seriousness

I'd like you to rate each medical visit on a scale from minor to serious. Tell me the scale number that best describes the medical visit. **[WRITE THE NUMBER ON THE CARD IN THE BOTTOM CORNER TRIANGLE. GO THROUGH EACH CARD, KEEPING A COPY OF THE SCALE IN FRONT OF THE RESPONDENT.]**

1	2	3	4	5
Very Minor	Minor	Moderate	Serious	Very Serious

[If respondent asks for interpretation, tell them the scale is open to their interpretation; they can use whatever criteria they consider appropriate.]

Rate Health

[REFER TO SUBJECT INFORMATION SHEET.]

Okay, now I would like you to evaluate your overall health. Would you describe your health as . . .? **[Get order of choices from subject information sheet.]**

(excellent, very good, good, fair, or poor)
(poor, fair, good, very good, or excellent)

Explain Seriousness

The last thing I'd like you to do, now, is describe how you decided whether an event was serious or minor. What makes an event serious? Or minor? **[PROBE TO GET THE DIFFERENCE IN A WAY YOU COULD EXPLAIN.]**

Mark Ending Time of Interview

[WRITE TIME, AND CLOSE UP.]

[END RECORDING.]

Appendix IV
Record matching and scoring

Contents and quality of medical records

Before the medical records were abstracted for this study, the George Washington University Health Plan's Quality Assurance Program conducted an audit to assess the accuracy of medical record documentation for primary care visits to the health plan. Receptionists' records of patients actually coming to the plan's offices were used as the criterion against which medical records were checked.

A stratified sampling plan representative of each day of the week, time of day, and provider was used to sample patient appointments during October and November of 1985. The medical record of each patient who had been checked in by the receptionist was reviewed by a research assistant to determine whether there was a record of the visit in the medical record. Additional information included the type of visit (acute, acute followup, chronic, health maintenance, and so forth) and whether the patient had a medical condition that could be considered chronic.

Results of an audit of 176 records showed that 169 out of 174 visits recorded by receptionists (97 percent) were documented in the patient's medical record. Consequently, the medical record information was judged acceptable for validating the accuracy of subject recall.

Record matching procedures

A physician's assistant on the health plan staff prepared abstracts of all records for all subjects. These abstracts contained information on event date, location (for example, office, home, emergency room), type (for example, primary care, specialist), purpose category (for example, history and physical, urgent, followup to urgent), provider type (for example, physician, nurse practitioner), provider name, specific reason, and treatment (prescription, tests, x ray, or referral). Meanwhile, researchers from the Human Resources Research Organization abstracted the same types of information from the subjects' information cards completed during the interview session.

One of the major methodological issues facing us was the development of criteria for determining that a match existed between a recalled event and a medical record entry. Before comparing the subjects' medical records with the recall abstracts, project researchers discussed the various types of information described above and how seriously they would evaluate each in determining whether or not two described visits really were the same event. During the course of these discussions, it became apparent that some types of information are regarded as more telling in making this determination than are others and that the existence of a match between the record and recall reports, the presence of contradictory information, and the lack of information in recall tend to be weighed differentially. The scoring system shown in table VI was agreed on.

As shown in table VI, this weighting scheme gives primacy to the visit reason in determining matches. Given that more than one event occurred for the same reason, similarity of dates weighs heavily in determining which record events are paired with which recalled events (this is particularly true in practice because such events are likely to all involve the same location, provider, and treatment).

Choosing a criterion for a match between the medical record entry and a recalled event was complicated by the fact that the number of points possible for event recall varied, depending on whether the medical record showed x rays, prescriptions, multiple providers, and so forth. Accordingly, for each event in the medical record, the number of points possible was computed, and the number of points for each record-recall event pair was converted to a proportion of the number of possible points for that medical event (possible points ranged from 55 to 85).

The percent criterion for scoring a "match" was derived after having researchers study a set of 20 possible record-recall event matches and independently rate each pair as "match" or "no match." The event pairs were then examined to see what criterion would maximize the number of match decisions that were the same as those made by the researcher after inspecting individual records and protocol extracts. For those pairs considered matches, point proportions ranged from 45 to 80 percent for one researcher and from 50 to 80 percent for the other, with an optimal cutoff score of 50 for the first researcher and 57 for the second. Based on their intuitive policies, a criterion of 53 percent was chosen for use in final recall scoring.

A procedure was then developed for identifying matches between medical records and recall events. For a given subject, each recalled event was compared to each record event and a "goodness of match" score assigned to the pair (using the point values shown in table VI). Thus, for a subject who had seven events on her medical record and who recalled six events, 42 record-recall event

Table VI. Scoring for matching medical record and recalled events

Information type	*Match*	*Mismatch*	*Only in medical record*
Reason[1] for event	+ 30 close match + 20 approximate	−20	([2])
Location (office visit, emergency room, and so forth)	+ 5	−10	–
Medical team (primary, specialty)	+ 5	−5	–
Provider type (MD, nurse, and so forth)	+ 5 each	–	–
Provider name	+ 10	–	–
Treatment, medicine, or referral	+ 5 each	−5 each	−3 each
Tests—x ray	+ 5 each	−5 each	−3 each
Date	+ 15 if ± 5 days + 10 if ± 15 days + 5 if ± 60 days	− 5 if > 60 days	–

[1]For visits dealing with multiple conditions, the subject had to recall all reasons to receive points for a close match on reason for event; subjects who remembered a subset of the conditions dealt with during the visit received points for an approximate match.
[2]"Don't know" responses were not allowed for this type of information.

"goodness of match" scores (all possible pairs) were computed. Each score, expressed as a proportion of the possible points for that record event, was then entered into a matrix of all record events crossed with all recall events. To determine matches, the researcher examined this matrix and identified the record-recall event pair with the highest "goodness of match" score. If this score exceeded the match criterion of 53 percent, the pair was ruled a match and all other pairs involving either that particular record event or that recalled event were eliminated from further consideration. The researcher then examined remaining matrix cells (record-recall event pairs) to identify the next highest "goodness of match" score, repeating the process described above until no remaining record-recall event pair had a "goodness of match" score exceeding the criterion (53 percent).

In addition, a rudimentary sensitivity analysis was performed by trying out an alternative criterion of 60 percent of possible points to assess the impact of the scoring criterion on the pattern of findings. The stricter scoring criterion did not alter the pattern of major findings: With the stricter scoring criterion, pre-postassessment group subjects recalled a mean of 34 percent of their medical record events before the intervention and 51 percent afterward. After the intervention, the recall performance of the postassessment-only group equaled that of the pre-postassessment group, just as with the 53 percent criterion. Chronic condition events were still remembered most poorly during free recall and benefited most from the intervention. Thus, the findings reported in the body of this report appear to be robust over moderate changes in scoring criteria.

Vital and Health Statistics series descriptions

SERIES 1. **Programs and Collection Procedures**—Reports describing the general programs of the National Center for Health Statistics and its offices and divisions and the data collection methods used. They also include definitions and other material necessary for understanding the data.

SERIES 2. **Data Evaluation and Methods Research**—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Studies also include comparison of U.S. methodology with those of other countries.

SERIES 3. **Analytical and Epidemiological Studies**—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

SERIES 4. **Documents and Committee Reports**—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

SERIES 5. **Comparative International Vital and Health Statistics Reports**—Analytical and descriptive reports comparing U.S. vital and health statistics with those of other countries.

SERIES 6. **Cognition and Survey Measurement**—Reports from the National Laboratory for Collaborative Research in Cognition and Survey Measurement using methods of cognitive science to design, evaluate, and test survey instruments.

SERIES 10. **Data From the National Health Interview Survey**—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in the continuing national household interview survey.

SERIES 11. **Data From the National Health Examination Survey and the National Health and Nutrition Examination Survey**—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

SERIES 12. **Data From the Institutionalized Population Surveys**—Discontinued in 1975. Reports from these surveys are included in Series 13.

SERIES 13. **Data on Health Resources Utilization**—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

SERIES 14. **Data on Health Resources: Manpower and Facilities**—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

SERIES 15. **Data From Special Surveys**—Statistics on health and health-related topics collected in special surveys that are not a part of the continuing data systems of the National Center for Health Statistics.

SERIES 16. **Compilations of Advance Data From Vital and Health Statistics**—These reports provide early release of data from the National Center for Health Statistics' health and demographic surveys. Many of these releases are followed by detailed reports in the Vital and Health Statistics Series.

SERIES 20. **Data on Mortality**—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.

SERIES 21. **Data on Natality, Marriage, and Divorce**—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.

SERIES 22. **Data From the National Mortality and Natality Surveys**—Discontinued in 1975. Reports from these sample surveys based on vital records are included in Series 20 and 21, respectively.

SERIES 23. **Data From the National Survey of Family Growth**—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a periodic survey of a nationwide probability sample of women 15–44 years of age.

SERIES 24. **Compilations of Data on Natality, Mortality, Marriage, Divorce, and Induced Terminations of Pregnancy**—Advance reports of births, deaths, marriages, and divorces are based on final data from the National Vital Statistics System and are published annually as supplements to the Monthly Vital Statistics Report (MVSR). These reports are followed by the publication of detailed data in Vital Statistics of the United States annual volumes. Other reports including induced terminations of pregnancy issued periodically as supplements to the MVSR provide selected findings based on data from the National Vital Statistics System and may be followed by detailed reports in Vital and Health Statistics Series.

For answers to questions about this report or for a list of titles of reports published in these series, contact:

Scientific and Technical Information Branch
National Center for Health Statistics
Centers for Disease Control
Public Health Service
Hyattsville, Md. 20782

301–436–8500