Retrospective Adult Assessment of Childhood Psychopathology

Ronald C. Kessler
Daniel K. Mroczek
Robert F. Belli

No. 012
Retrospective Adult Assessment of Childhood Psychopathology

Ronald C. Kessler, Ph.D.
Daniel K. Mroczek, Ph.D.
Robert F. Belli, Ph.D.

Institute for Social Research
University of Michigan

March, 1994 (Revised)
This chapter discusses methodological issues involved in obtaining accurate retrospective reports about childhood psychopathology from epidemiologic surveys of adults. It is obvious that such data would be subject to greater recall failure than prospective reports obtained from the same respondents by interviewing them at regular intervals during their childhood. However, a study in which a large general population sample of children is interviewed throughout their childhood would be both enormously expensive and fraught with logistic challenges. This could explain why such a study has never been carried out. Instead, our understanding of the developmental features of childhood psychopathology is based on the results of more limited investigations, consisting for the most part of small in-depth longitudinal studies of high-risk children (e.g., Block, Gjerde, & Block, 1991; Kovacs & Paulauskas, 1984) or children in treatment (e.g., Harrington, Fudge, Rutter, Pickle, & Hill, 1990), and surveys of children and adolescents in the general population (e.g., Offord, 1985; Cohen, Cohen, & Brook, 1993; Cohen, Cohen, Kasen, & Velez, 1993). Valid retrospective studies about childhood psychopathology obtained from adult respondents would supplement these existing studies and help identify the adult outcomes associated with childhood disorders.

Very little epidemiologic research has been conducted using retrospective adult reports of childhood psychopathology; we are unaware of any research that has attempted to validate such reports. However, a considerable body of research has used retrospective reports from adults to study the long-term effects of childhood adversities on adult psychopathology (e.g., Harris, Brown, & Bischof, 1990; Holmes & Robins, 1988; Kessler & Magee, 1993). A review of the methodological evaluation of these studies (e.g., Christianson & Loftus, 1991; Gotlib, Mount, Cordy, & Whiffen, 1988; Robins et al., 1985) may suggest techniques for obtaining valid information about childhood psychopathology.

One of the most consistent results of these studies is that adults generally have no direct recollection of experiences that occurred during the first five years of their lives (Wetzler & Sweeney, 1986). This is consistent with neurophysiological studies showing that the corpus callosum is incompletely myelinated prior to age six (Baddeley, 1986). Based on these results, it seems unlikely that retrospective reports from adults will provide accurate information about childhood psychiatric disorders prior to age six.

Adult retrospective reports about childhood experiences at age six or later, however, are apparently much more reliable. A recent review of methodological studies concludes that "adults asked to recall salient factual details of their own childhoods are generally accurate, especially concerning experiences that fulfill the criteria of being unique, consequential, and unexpected" (Brewin, Andrews, & Gotlib, 1993). Two important caveats must be noted: First, the evidence is clear that people cannot remember the incidental details of childhood experiences, only the central features (Sheingold & Tenney, 1982). Second, the evidence is equally clear that accurate recall requires the respondent to engage in active and extensive memory search. To ensure this precision, the respondent must understand the task must be motivated to engage in hard work rather than supply a superficial response (Cannell, 1985a, 1985b; Cannell, Oksenberg, & Converse, 1977; Hippler & Schwarz, 1986). Unfortunately, many respondents are prone to draw lay theories of childhood development and current self-images or self-knowledge to infer their childhood attributes rather than attempt to recall them (Ross, 1989). While this theory-guided self-inference requires much less work than active memory search, it usually yields less accurate retrospective reports (Ross & Conway, 1986).

As the few extant studies of childhood psychopathology have not been validated, it is unclear whether early psychiatric symptoms could be recalled as accurately as other childhood experiences. The methodological results reviewed above suggest that researchers must (1) impress on respondents that active memory search is required, (2) motivate active memory search and honest reporting, (3) recognize the limits of autobiographical memory by limiting questions to those which can be answered accurately, and (4) develop procedures appropriate for use in general population surveys to facilitate active memory search.

A considerable amount of methodological research has been carried out by survey researchers on each of these four topics (e.g., Moss & Goldstein, 1979; Bradburn, Sudman & Associates, 1979; Cannell, Miller, & Oksenberg, 1981). This research has advanced considerably over the past decade as cognitive psychologists have become interested in the survey interview as a natural laboratory for studying memory processes (e.g., Biderman, 1980; Jabine, Straf, Tanur, & Tourangeau, 1984; Tanur, 1992). A number of important insights have emerged from this work that suggest practical ways of improving the accuracy of retrospective reports. We review these insights in the present chapter and discuss modifications which were made to a standardized psychiatric diagnostic interview for use in the National Comorbidity Survey (Kessler et al., 1994) to enhance recall of childhood psychopathology.
1. The Importance of Understanding the Task

Obviously ambiguous questions will likely be misconstrued. Less obvious, though, is just how ambiguous questions are posed in standard epidemiologic surveys and how often respondents must “read between the lines.” Belson (1981) investigated this issue in a set of standard survey questions and found that more than 70% of respondents interpreted at least some questions differently from the researcher, leading Belson to conclude that subtle misinterpretations are pervasive in survey situations. Oksenberg, Cannell, and Kalton (1991) came to a similar conclusion in their debriefing of a nationally representative sample of respondents who were administered standard health interview survey questions. At least one key phrase in two-thirds of the questions in their analyses were misinterpreted by respondents. Both Belson and Oksenberg et al. found, furthermore, that respondents generally believed that they understood what the investigator meant even when their interpretations of the questions were quite idiosyncratic.

How is it possible for there to be so much misunderstanding? As Oksenberg and her colleagues discovered, the answer lies partly in the fact that such terms as “physical examination,” “stay in bed,” and “doctor” have different meanings for different people. But beyond these rather obvious examples of vaguely defined terms is the more fundamental fact that the survey interview situation is a special kind of interaction in which the standard rules of conversation -- rules which help fill in the gaps in meaning that exist in most speech -- do not apply. Unlike the situation in normal conversational practice, the respondent in the survey interview often has only a vague notion of to whom he is talking or the purpose of the conversation (Cannell, Fowler, & Marquis, 1968). The person who asks the questions (the interviewer) is not the person who formulated the questions (the researcher), and the questioner is often unable to clarify the respondent’s uncertainties about the intent of the questions. Furthermore, the flow of questions in the interview is established prior to the beginning of the conversation, which means that normal conversational rules of give and take in question and answer sequences do not apply. This inflexibility leads to such odd interactional moments as when the interviewer asks a question which the respondent previously answered while elaborating an earlier response and when the interviewer answers a query about the meaning of a question by telling the respondent means “whatever it means to you.” These out-of-the-ordinary interactions occasion more misreading than do normal conversations even when questions are seemingly straightforward (Clark & Schober, 1992).

This problem of meaning was noted many years ago by Lazarsfeld (1934), who demonstrated that a question as simple as “Why did you buy this book?” will be interpreted in vastly different ways depending on the respondent’s understanding of whether the interviewer is concerned with why YOU bought the book (as opposed to your wife buying it for you), why you BOUGHT this book (as opposed to borrowed it), why you bought THIS book (as opposed to some other book), or why you bought this BOOK (as opposed to buying something else with the money). The respondents’ understanding of which connotation the interviewer intends is often determined by subtleties such as voice inflection and contextual issues.

Our own pilot work for the National Comorbidity Survey (NCS) found that misinterpretation is especially common with the inquiries about psychiatric disorder that appear in the original version of the Composite International Diagnostic Interview (CIDI; Robins, Wing, Wittchen, & Helzer, 1988), the structured diagnostic interview used in the NCS. There are a variety of reasons for this -- including ambiguous question wordings and awkward transitions -- but our debriefing revealed that substantial confusion arose from respondents’ failure to understand the purpose of the questions. For example, a substantial number of respondents misinterpreted the intent of such recall questions as “In your lifetime, have you ever had two weeks or more when nearly every day you felt sad, blue, depressed?” The misinterpretation concerned the task itself. Only about half of pilot respondents interpreted the question as it was intended by the authors of the CIDI; namely, as a request to engage in active memory search and report episodes of the sort in the question. The other respondents interpreted the question as a request to report whether a memory of such an episode was readily accessible. These latter respondents did not believe that they were being asked to engage in active memory search and did not do so. Not surprisingly, these respondents were much less likely than those who understood the intent of the question to remember lifetime episodes.

Why did so many respondents misinterpret the intent of these lifetime recall questions? As Marquis and Cannell (1969) discovered in their research on standard interview practice, respondents are generally ill-informed or poorly motivated. Lacking clear instructions and having little grasp of research aims, they render only desultory answers. Debriefing indicates that most respondents consider being interviewed “a lark -- an unimportant and uninvolving activity (Cannell et al., 1981).” As noted by Clark and Schober (1992) in their analysis of discourse rules in survey interviews, the interaction flow in most surveys reinforces the perception that careful response is unimportant. Normal rules of
conversation require a person who is asked a question signify recognition of turn-taking either by answering the question or by making some other relevant comment (such as, "um, let me see now...") within about one second after the question is issued (Jefferson, 1989) unless there is an explicit instruction on the part of the questioner to the contrary. When an interviewer asks a question that requires considerable thought, the respondent is likely to assume in the absence of instructions to the contrary that the interviewer is operating under normal conversational rules and, as such, is really asking for an immediate and appropriate answer.

The work of Cannell et al., (1981) shows that this conversational artifact can be minimized by explicitly instructing respondents to answer completely and accurately. The use of such instructions can substantially improve the quality of data obtained in surveys. Our pilot work for the NCS built on this result by investigating the effect of adding clarifying statements throughout the clinical interview aimed at informing respondents that accuracy was important. For example, we experimented with the following introduction to CIDI stem questions for lifetime recall of specific psychiatric disorders: "The next question might be difficult to answer because you need to think back over your entire life. Please take your time and think carefully before answering."

The use of this introduction prompted respondents to consider longer whether they ever had such experiences as "two weeks or more when nearly every day you felt sad, blue, depressed" or "a period of a month or more when most of the time you felt worried or anxious." Debriefing showed that this slower response time was attributed to the fact that respondents were engaging in active memory search rather than estimating. A subsequent question wording experiment showed that the use of this introduction led to a significant increase in the proportion of respondents in a national sample who endorsed lifetime diagnostic stem questions for a variety of affective and anxiety disorders.

2. The Importance of Motivation

One problem with emphasizing to respondents the need to work hard at a series of demanding and potentially embarrassing recall tasks is that more respondents than otherwise may refuse the job. Recognition of this problem among survey research methodologists has led to the development of motivational techniques which are intended to increase the chances that respondents will accept the job of answering completely and accurately. Three techniques that have proven to be particularly useful in this regard are the use of motivational components in instructions, the use of contingent reinforcement strategies embedded in interviewer feedback probes, and the use of respondent commitment questions.

2.1. Motivational instructions

There is evidence that clarifying instructions and research aims can help motivate complete and accurate reporting (Cannell et al., 1981). Debriefing shows that respondents are more willing to understand laborious and possible painful memory searches if they recognize some altruistic benefit of doing so. Even such an uncompelling rationale as "it is important for our research that you take your time and think carefully before answering" has motivational force. This is even more so when instructions include statements that have universalistic appeal, such as: "accuracy is important because social policy makers will be using these results to make decisions that affect the lives of all of us."

Second, instructions that define the nature of interviewer expectations for respondent behavior help to establish a perspective on the interview that can have motivational force. The literature on cognitive factors in surveys contains many examples of the subtle ways in which perspectives established in questions subsequently influence respondent behaviors. For example, Loftus and Palmer (1974) showed respondents a film of an automobile accident and asked them to estimate the speed the cars were travelling prior to the accident. Respondents estimated the rates as significantly greater if they were asked how fast the cars were travelling "when they collided" rather than "when they contacted each other." This same literature shows that perspective can have motivational force when it implies a common purpose (Clark & Schober, 1992). That is, if a question is posed in such a way that it implies that hard work will be invested in arriving at an answer, it is incumbent on the respondent either to demure explicitly or tacitly to accept the task of working hard as part of the common understanding between interviewer and respondent. By answering the question, the respondent, in effect, makes a commitment to honor the injunction implied in the perspective of the question and this implied commitment, in turn, creates motivation to this task (Marlatt, 1972).
2.2. Contingent reinforcements

Consistent with research on behavioral modification of verbal productions through reinforcement (e.g., Centers, 1964), several survey researchers have demonstrated that verbal reinforcers such as "thanks" and "that's useful" can significantly affect the behavior of survey respondents. Marquis and Cannell (1969), for example, showed experimentally that the use of such reinforcers resulted in a significant increase in the number of chronic conditions reported in response to an open-ended question about illnesses. These feedback remarks are often used in an unsystematic way, however, as part of general procedures to build and maintain rapport rather than in a systematic way to reinforce good respondent performance.

Based on these observations, Cannell and his associates developed a method for training interviewers to use systematic feedback -- both positive and negative -- to reinforce respondent effort in reporting (Oksenberg, Vinokur, & Cannell, 1979a). The central feature of this method is the use of structured feedback statements coordinated with the content and timing of instructions aimed at reinforcing respondent performance. It is important to recognize that it is performance that is being reinforced rather than the content of particular answers. For example, a difficult recall question may be prefaced with the instruction "This next question may be difficult, so please take your time before answering." In contingency feedback instruction interviewers would issue some expression of gratitude whenever the respondent seems to consider his or her answer carefully, whether they remembered anything or not. Alternatively, the interviewer might instruct the precipitous respondent: "You answered that awfully quickly. Was there anything (else), even something small?" Such invitations to reconsider would occur whenever the respondent gave an immediate answer whether or not anything was reported. This structured feedback is programmed periodically throughout the interview in order to maintain the focus on performance standards and to reinforce motivation.

Experiments carried out by Cannell and his associates (Miller & Cannell, 1977, Vinokur, Oksenberg, & Cannell, 1979) have documented that the combined use of these contingent reinforcement probes with instructions explaining the importance of careful and accurate reporting leads to substantial improvement in recall of health-related events in general population surveys, including validated dates of medical events. Importantly, their results also show that self-enhancing response biases are reduced when these strategies are used, as indicated by both a decreased tendency to under-report potentially embarrassing conditions and behaviors (e.g., gynecologic problems, seeing an X-rated movie) and a decreased tendency to over-report self-enhancing behaviors (e.g., number of books read in the last three months, read the editorial page of the newspaper the previous day).

2.3. Commitment questions

We noted above that instructions often have the effect of eliciting indirect commitment to the goal of serious and complete reporting. It is also possible to motivate the respondent to accept this goal by asking an explicit commitment question as part of the interview. We did this in the NCS by prefacing the section of the interview that asked a series of lifetime diagnostic stem questions with the following commitment question:

This interview asks about your physical and emotional well-being and about areas of your life that could affect your physical and emotional well-being. It is important for us to get accurate information. In order to do this, you will need to think carefully before answering the following questions. Are you willing to do this?

Consistent with the results of previous studies using similar questions (Cannell et al., 1981), we found that only a small fraction of respondents answered negatively (only 35 of the 8133 people who began the interview). These interviews were terminated, based on the decision not to invest interviewer time on respondents not willing to work seriously at the task.

Experimental studies carried out by Cannell and his associates (Cannell et al., 1981; Oksenberg et al., 1979a, 1979b) have shown that commitment questions improve accuracy of recall. Furthermore, their studies indicate that the joint use of motivating instructions, contingency feedback, and a commitment question has an interactive effect that increases the intensity of memory search and accuracy beyond the effects of any one component separately. This extends not only to the proportion of respondents in different experimental conditions who recall and report past experiences, but also to other indicators of commitment such as amount of detail reported and use of personal records and other outside information sources as memory aids during the course of the interview.
3. The Limits of Autobiographical Memory

3.1. Episodic and semantic memories

Research on basic cognitive processes has shown that memories are organized and stored in structured sets of information packages commonly called schemas (Markus & Zajonc, 1985). When the respondent has a history of many instances of the same experience which cannot be discriminated, the separate instances tend to blend together in memory to form a special kind of memory schema called a "semantic memory," a general memory for a prototypical experience (Jobe, White, Kelley, Mingay, Sanchez, & Loftus, 1990; Means & Loftus, 1991). For example, the person may have a semantic memory of what panic attacks are like but, due to the fact that he has had many such attacks in his lifetime, cannot specify details of any particular panic attack. In comparison, when the respondent has had only a small number of lifetime experiences of a certain sort or when one instance stands out in memory as much different from the others, a memory can likely be recovered for that particular episode. This is called an "episodic memory."

In the case of memories of illness experiences, memory schemas tend to include not only semantic memories of prototypic symptoms but also personal theories about causes, course, and cure (Leventhal, Nerenz, & Steele, 1984; Skelton & Croyle, 1991). Some of these theories will conceptualize the experience in illness terms and others as a moral failing, a punishment from God, or a normal reaction to stress (Gilman, 1988). These interpretations influence the extent to which different memory cues are capable of triggering the schemas.

The effects of memory schemas and the difference between semantic and episodic memories are central themes in research on autobiographical memory. Indeed, we must determine whether episodic memories can be recovered and whether the respondent is answering the questions by referring to episodic memories or by drawing inferences of what the past must have been like on the basis of more general semantic memories. Research shows that people are more likely to recover episodic memories for experiences that are recent, distinctive, and unique, while for experiences that are frequent, typical, and regular, people will rely more on semantic memories (Belli, 1988, Brewer, 1986; Menon, 1994).

3.2. Asking questions without knowing the limits of memory

When a survey question is designed to ask about a particular instance of an experience, it must be posed in such a way that the respondent knows he or she is being asked to recover an episodic memory. The researcher must have some basis for assuming that an episodic memory can be recovered for this experience. If it cannot, a question that asks for such a memory implicitly invites the respondent to infer or estimate rather than remember and this can have adverse effects on quality of reporting later in the interview (Pearson, Ross, & Dawes, 1992). In comparison, when a question is designed to recover a semantic memory or to use semantic memories to arrive at an answer by estimation, that should be made clear.

One difficulty with these injunctions in the case of retrospective recall questions about lifetime psychiatric disorder is uncertainty about what level of recall accuracy to expect. We confronted this problem in pilot studies for the NCS when we asked the standard CIDI questions about first onset, such as the panic onset question "When was the first time you had one of these sudden spells of feeling frightened or anxious and had these problems like (PREVIOUSLY ENDORSED SYMPTOMS OF PANIC)?" Debriefing of pilot respondents revealed that some people had very vivid memories of their first panic attack, while others had no such memory. The problem posed by this variation was how to develop a method of asking the question that reinforced our overall commitment to collecting complete and accurate information, while simultaneously recognizing the limits of autobiographical memory and avoiding a request for a precise answer from the subsample of respondents who were unable to recover an episodic memory for their first episode.

We resolved this problem by adapting several of the principles discussed above to a three-part question series designed to inform respondents that answers should be as precise as possible while still recognizing the limits of memory. The question sequence began with what has been referred to in the literature as a "prequest," a question aimed at clarifying the nature of the request for information in subsequent questions. The prequest question was:

Can you remember your EXACT age the VERY FIRST TIME you had a sudden spell of feeling frightened or anxious and had several of these other things ('other things" refers to a checklist of symptoms that respondents previously reported which was presented for visual review on a cue card) at the same time? (EMPHASIS IN ORIGINAL)

During the pilot work we probed positive responses to determine the basis for exact recall and discovered that, in
3.3. Increasing knowledge about the limits of memory

As the discussion in the last few paragraphs makes clear, a major barrier to evaluating strategies for improving recall of psychiatric disorders is that we lack any clear understanding of the limits of autobiographical memory for this class of experiences. This is true, in large part, because we lack a clear validation standard. The same limitation plagues much of the research on long-term autobiographical memory in surveys. Although a small number of record-check studies have been carried out to validate the accuracy of retrospective reports about such things as hospitalizations (Marquis, Cannell, & Laurent, 1972), doctor visits (Means & Loftus, 1991), voting behavior (Abelson & Loftus, 1992), and income tax returns (Withey, 1954), these are exceptions. Most methodological studies of memory in surveys are of two other sorts. One of these uses short-term test-retest designs in which respondent reports of attitudes or occasionally of recurring behaviors like dietary intake (Smith, Jobe, & Mingay, 1991a; 1991b) in a baseline survey are used as the validation standard. Recall bias is then studied in a retest interview administered a few hours, a few days, or a few weeks later, often with experimental manipulations designed to assess the extent to which recall can be influenced by various types of recall cues (e.g., Pearson et al., 1992). The other type of design uses experimental manipulations such as the feedback and commitment probes described above to study changes in retrospective reports about more distant events, but without any independent source of data on the actual occurrence of these events as a validation standard (e.g., Crespi & Swinehart, 1982). Based on these two types of studies, we know that there are substantial errors in short-term retrospective reports of recurring behaviors ("what did you have to eat last Thursday?") and biases in the direction of reporting more consistency than really exists in recall of attitudes after an experimental attitude change manipulation. These studies provide no clear evidence, however, concerning the limits of autobiographical memory for the sort of provocative, but perhaps chronologically distant experiences targeted in retrospective studies of childhood psychopathology; namely, reports about such things as recurrent depression, substance use problems, unreasonable fears, and behavior problems that occurred during one's childhood. As a result of this limitation, we are uncertain about the real limits of autobiographical memory for childhood psychopathology.

Naturalistic and experimental studies to obtain this type of data are needed. This could be done using the model of previous validation studies of health experiences (e.g., Loftus, Smith, Klinger, & Fielder, 1992; Marquis et al., 1972; Means & Loftus, 1991). Using medical records, one could select a sample of people known to have been in treatment in the past and assess the extent to which they report this experience in a household survey in which they are initially blind to the fact that their medical records are known. Concerns could be raised, however, that this type of study might overestimate the accuracy of long-term recall of childhood disorder because children who were in treatment are unrepresentative of all children with psychiatric disorders and are probably more aware of their disorders and have distinctive memory cues associated with the treatment experience that could facilitate long-term retrospective recall. Based on these concerns, a more valid type of study would be one in which data concerning childhood disorders derive
from some source other than treatment records. An obvious possibility is to reinterview a subsample of adult respondents who participated during their childhoods in a general population survey of mental health and to use the original survey data as a validation standard in evaluating the accuracy of retrospective reports.

As noted in the introduction, long-term retrospective validation studies of this type have been carried out to study the accuracy of retrospective reports about other aspects of childhood. Robins and her colleagues (1985), for example, evaluated the accuracy of retrospective reports about the childhood home environment in a sample of adults who had attended a child guidance clinic when they were children. The study compared the recollecting of respondents in their 40s with data obtained thirty years prior. Subjects were able to provide quite good reports about major aspects of the family environment (e.g., parent work histories and family breakups) as well as specify objective features of family life that did not involve value judgments (e.g., whether the parents fought in front of the children or hit each other). Reports were much less accurate when the questions involved judgments or interpretations rather than factual descriptions. These results allowed Robins to characterize the limits of adult autobiographical memory concerning the childhood home environment and to develop a Home Environment Interview that focused on concrete descriptive questions which could be recalled with good accuracy. Precisely this sort of work is needed to extend our understanding of the limits of adult autobiographical memory for childhood psychiatric disorders and to develop an informed strategy for stimulating adult recall of these disorders.

Ideally, such validation studies should include both a discovery and an experimental component in addition to the investigation of variation in the accuracy of different types of recall questions. The discovery component should follow the procedures used in recent inductive cognitive psychological studies of memory processes (e.g., Oksenberg et al., 1991; Fisher & Geiselman, in press) to elicit information from respondents about the memory search strategies used in an effort to discover whether some search strategies are superior in recovering certain types of memories. If so, this information could be used in the development of subsequent interview protocols to coach respondents in the use of effective strategies. For example, childhood psychopathology might be more easily recalled by respondents who begin thinking about their earliest memories and then move forward in time rather than begin with recent memories and move backward in time (Fathi, Schooler, & Loftus, 1984; Loftus & Fathi, 1985). If this is verified, an instruction could be given to all respondents to use the forward search procedure. The experimental component should build on the results of the discovery component as well as on the results of the accumulated literature to develop a series of nested experiments to evaluate the effects of various interventions on accuracy of recall.

It is important to embed experiments in recall validation studies as most techniques which improve the accuracy of autobiographical memory are known to fail in some situations. For example, Crespi and Swinchart (1982) developed a dual time frame approach to improve accuracy of reporting and dating a variety of recent (past two months) health-related actions (e.g., blood pressure check, eye examination). A subsequent validation study (Loftus et al., 1992) showed that this method does, in fact, improve accuracy of reports about this class of behaviors. However, other researchers (Abelson & Loftus, 1992) found that the same strategy was ineffective in improving accuracy of self-reported voting behavior. Results such as these indicate that the processes which influence accuracy of reporting vary by context and that the effectiveness of particular strategies must be reconsidered whenever a new domain of memory is under investigation.

4. Facilitating Complete and Accurate Recall and Reporting

We reviewed several strategies that optimize recall accuracy, including explicit instructions respondents that complete and accurate answers are expected, as well as various techniques to motivate the respondent to provide such answers. Other techniques, however, have been developed to improve accuracy once respondents commit to active memory search. These latter techniques provide recall aids that increase the efficiency of memory work. Again, the technique used depends on the type of memory targeted. The following section considers how to facilitate memory in the context of four broadly conceived questions concerning recall of childhood psychiatric disorders. There are, of course, more than four questions of interest concerning such disorders, but these four are central and raise a number of important issues that have been addressed in the literature. The four questions are: "Did such a disorder ever occur?" "What were the associated symptoms and role impairments?" "How old were you the first time this happened?" and "What was the course of the disorder after that first occurrence?"
4.1. Remembering whether a disorder ever occurred

A good deal of evidence suggests that people who have never had certain experiences quite accurately answer no to "Have you ever ...?" questions (Glucksberg & McElrath, 1981; Shannon, 1979). A core process used in making these judgments is the lack-of-knowledge inference (Genter & Collins, 1981), the conclusion drawn from lack of knowledge of an experience that it ever happened to me. For example, Lessler, Salter, & Tourangeau (1989) reported that survey respondents who had never heard of dental sealants felt quite confident in saying that they had never worn such sealants. The potential problem here is that the way in which the experience is characterized can importantly impact the perception of lack-of-knowledge. This is an especially serious issue in lifetime diagnostic stem questions, which often deal with vaguely defined terms that are easily misconstrued, to prompting quick and incorrect lack-of-knowledge inference. We discuss below several techniques that can be used to minimize this potential problem.

4.1.1. The pace of the interview. A number of survey methodologists have noted that unless interviewers are carefully trained to the contrary, they will ask questions too quickly and that this will reduce the accuracy of respondent reports (Cannell et al., 1977; Neter & Waksberg, 1964; Sudman & Bradburn, 1982). This is especially true for lifetime recall questions. At least two fairly obvious processes are involved here. Haste on the part of the interviewer conveys the message that quick response is more important than accurate response (Clark & Shober, 1992). Also, memories are more likely to be recovered when respondents are allowed to think at their own pace rather than rushed (Bradburn, Rips, & Shevell, 1987). Based on these observations and on the analysis of interaction sequences in interviews, Cannell and his associates have recommended that interviewer reading pace should be no more than an average of two words per second (Cannell et al., 1981), that respondents should be explicitly asked to think at their own pace (Cannell & Kaha, 1968), and that critical questions should be designed to encourage periods of silence that are explicitly defined as thinking time (Cannell, 1985a). Several experiments have documented that these procedures lead to more accurate recall of health-related events (Burton & Blair, 1991; Means, Swan, Jobe, & Esposito, 1993; Lessler et al., 1989).

4.1.2. Memory cues. Two general types of memory cues that have been used by researchers to stimulate recall of past life events might also assist relevant recall of lifetime psychiatric disorder: concrete cues and context cues. Concrete cues consist of very explicit questions or other stimuli (pictures, smells, sounds) that are aimed at triggering memories for particular experiences. Surveys about the use of medications, for example, sometimes use multicolor pill cards to stimulate memory of particular medications that the respondent has taken (Parry, Balter, & Cisin, 1970-1971). Other surveys use such recall aids as lists of products to stimulate recall of purchases or lists of life events to stimulate recall of stressful experiences. A few surveys about particularly complex topics such as hospital visits or expenses have even requested respondents -- by means of an advance letter -- to review their records and have them available as memory cues during the interview (Sudman & Bradburn, 1982). Experimental evidence shows that these strategies do, in fact, improve recall (Bradburn et al., 1987). Cues of this sort might be especially useful in asking lifetime diagnostic stem questions by including very concrete characterizations that retrieve relevant memory schemas.

Context cues consist of broader questions about a particular life domain that facilitate recall of events within that domain by triggering a common memory schema which can be used to structure information search (Higgins, Rholes, & Jones, 1977). For example, comparative studies show that life event reports are increased when questions about events relating to a particular area of life are embedded in a larger series of survey questions about that area (Kessler & Wethington, 1991). Context cues also appear to have a significant effect on lifetime recall of psychiatric disorders, as documented in a recently completed experiment in which we administered a series of diagnostic stem questions to a nationally representative sample of adults in a telephone survey. A random half of the respondents were presented with an introductory statement about the importance of careful and complete reporting and then administered a commitment question prior to a series of eight diagnostic stem questions about the lifetime occurrence of such experiences as a period of two weeks of feeling sad or blue, a period of one month or more of feeling worried or anxious, and a sudden spell or attack of panic. Remaining respondents were administered the same introduction and commitment questions, but then received ten context questions about life "when you were a child up to the age of 12." The purpose of these questions was to provide a memory context for that part of the life span by asking such things as the following:

During those years, did you live in just one place, move once or twice, or move around a lot? How do you think (staying in the same place/moving/moving around a lot) affected your childhood? Overall, how was your relationship with your parents or the people who raised you during those years? Did you have brothers and sisters living at home with you during those years? (If YES, How well did you get along with them?)

IF NO,
How did you feel about being an only child?)

These questions were then followed by the same eight diagnostic stem questions administered to the first
half-sample, excepting they were confined to the time when the respondent was less than 12 rather than over their entire
lifetimes. This same sequence -- ten context questions followed by eight diagnostic stem questions -- was then repeated
for each of three other periods in the life course: "during your teenage years," "during your 20s," and "from the age of 30
up to the present." The data were then aggregated and compared to determine whether this procedure led to more
complete reporting. There was in fact a significant increase in the proportion of respondents who responded positively to
the diagnostic stem questions when they were embedded within the stage-of-life context questions. In addition,
follow-up questions that administered the full CIDI to all respondents who endorsed stem questions showed that not only
the stem questions but also the estimated lifetime prevalences of almost all the DSM-III-R disorders for which stem
questions were included in the battery were significantly higher in the stage-of-life context subsample than the control
subsample. Furthermore, respondents in the former subsample reported their ages of onset as being earlier and were
more likely to say that they could clearly remember the first episode of their disorders than respondents in the control
subsample.

4.1.3. Multiple frames of reference. We noted above that use of a lack-of-knowledge heuristic can lead to false
conclusions about an experience never happening if the terms used are inconsistent with the schema in which memory
for the experience is stored. This is a well-known phenomenon in life events research. For example, a mother who finds
that her teenage son has been stealing money from her purse is unlikely to report this in response to a question about
"robbery" because the experience is coded as something quite different even though it fits the legal definition of robbery.
In order to reveal this experience a different memory schema is required, one that asks about such things as having a
major disappointment with a close friend or relative, or finding out something about a person close to you that was very
upsetting, or being betrayed by a loved one. Life events researchers have shown that completeness of recall of life events
can be substantially improved by using questions of the latter sort to elicit multiple frames of reference that trigger
relevant memory schemas for events (Brown & Harris, 1978; Kessler & Wethington, 1991). The same insight is used in
studies of eyewitness memory for the details of crimes (Fisher, Geiselman, & Amador, 1989, Fisher & Geiselman, in
press; Fisher & Quigley, 1992), where one of the core principals of stimulating accurate recall is to probe multiple
representations of the situation.

This insight might be used in several ways to improve the accuracy of lifetime diagnostic stem questions regarding
mental disorders. The most powerful elaborates the descriptive details presented in the stem questions, recognizing that
these questions are, in effect, vignettes intended to describe a syndrome in such a way that triggers a memory schema. In
an effort to gain some insight into the ways this might be done most effectively, interviews should be conducted with
children who are experiencing disorders of particular types to elicit information about the ways they describe their
experiences. Such lay representations of personal illness experiences have been shown to share a number of core
features across people with the same conditions (Leventhal, Meyer, & Nerenz, 1980; Bishop, 1991; Bishop & Converse,
1986), and to vary in systematic ways across different segments of the population (Angel & Thoits, 1987; Gilman, 1988;
Kleinman, 1986). Knowledge of these schemas may assist the researcher in facilitating accurate recall of disorders by
improving the extent to which diagnostic stem questions correspond to the schemas.

The possible utility of this approach is illustrated by recent work with Black and Latino psychiatric patients to learn
how they describe their disorders. This was done in an effort to refine the diagnostic stem questions for affective
disorders and anxiety disorders in the CIDI. The interviews, carried out by medical anthropologists, asked respondents
to describe in their own words what it was like to have their problems. After doing this, respondents were presented with
the relevant diagnostic stem questions from the CIDI and asked if they would have endorsed these descriptions if they
had been administered as part of a survey. Respondents were then asked to rewrite the questions so they would more
closely match their own experiences.

The results led to several important insights about ways to expand the standard CIDI questions. For example, in the
case of panic disorder, the CIDI stem question is: "Have you ever had a spell or attack when all of a sudden you felt
frightened, anxious or very uneasy in situations when most people would not be afraid or anxious?" Yet the word
"attack" was not used by people when describing these panic events. Nor were these words used when respondents
rewrote the questions. Instead, respondents spoke of the episodes as simply occurring "all of a sudden." We asked about
this failure to use the terms "spell" and "attack" and discovered that many respondents felt that these words could be
confused with other things that were recognized as experiences that occurred in their cultural communities ("spell")
referring to magic and "attack" to violence). As a result, they did not want to use these words in their own descriptions. Based on these reports, it might make sense to remove these words from any revision of the CIDI stem question and emphasize rather the sudden nature of the episode.

We also discovered that when respondents spoke of "being frightened, anxious or very uneasy," they did so in much more evocative terms than the CIDI, describing times when they suddenly "had a wave of terrible fear" wash over them; when they experienced a sense of terror, when they felt as if they were "going to go crazy any minute," when they "totally lost control and wanted to scream," and when they "wanted to run away or escape but had no place to go." The phrase "very uneasy" virtually never arose in the rewritten stem questions because, according to respondent accounts, "uneasy" is too mild a word to describe their experience. Such observations provide a number of clues about ways to write a new series of panic stem questions, each emphasizing a somewhat different cognitive-emotional dimension of panic that may tap into an illness representation more accurately than the current stem question. The use of more evocative stem questions might plausibly lead to substantial improvement in recall, as there is evidence that emotionally-charged cognitions decay more slowly in memory and can be recovered with memory cues that emphasize the content of these emotions (Reisberg & Heuer, 1992).

We also found that physiological symptoms figured much more prominently in the scripts written by people with panic than in the CIDI stem question: not being able to breathe, having one's heart beat fast, feeling faint, and a perception of impending death were commonly cited by sufferers of panic. Based on these observations, it would seem wise to develop one or more variants on this panic stem question that emphasized physiological symptoms.

Related to these physiological descriptions was an objection on the part of quite a few respondents to the secondary clause in the CIDI stem question that the panic must occur "in situations when most people would not be afraid or anxious." The respondents who objected in this way reported that they would have denied the CIDI question if they had been in a survey because their intense fears only occur at times when their heart starts suddenly pounding and they get dizzy and they think they're going to die, a situation in which anyone would be afraid. Accordingly, panic stem questions might be revised to exclude mention of nonprovoking situations and rather, attempt to discriminate panic (attacks that occur unexpectedly and without a precipitating cause) from phobia in a separate series of questions administered after establishing the existence of periods of suddenly feeling very frightened.

It is unclear whether children would be able to provide as rich characterizations as those obtained in these interviews with adults. Nor is it clear that the descriptions supplied by children would usefully determine diagnostic stem questions for adults recalling childhood disorders; as we know, illness representations change over time as people gain more personal experience with their symptoms (Safer, Tharps, Jackson, & Leventhal, 1979; Leventhal & Diefenbach, 1991). Nonetheless, the above example illustrates that insights into designing more evocative questions can be obtained by offering people an open-ended opportunity to describe their own experiences. Although we are unaware of any work that has been done along these lines for childhood psychiatric disorders, we suspect that useful information would result by listening to children, adolescents, young adults describe their symptoms. Special notice should be taken of changes depending on time elapsed since the childhood experiences and/or progression of the disorder.

4.2. Remembering symptoms and role impairments

After eliciting a memory for a lifetime diagnostic stem question it is important to query about additional criteria needed to qualify for a diagnosis. Depending on the diagnosis under consideration, the respondent must describe the existence and duration of symptoms and role impairments as well as provide information regarding possible organicity. The demands on respondent memory imposed by these questions are quite different from those associated with answering diagnostic stem questions. In particular, symptom and impairment questions are asked in the context of a prior question that focuses memory on a specific time; an advantage over diagnostic stem questions. This focusing is known to have a positive effect on recall accuracy (Pearson et al., 1992; Hasher & Griffin, 1978), depending on whether the respondent can recall a particular instance of the disorder or only retrieves a semantic memory.

When episodic memories cannot be readily recalled, research suggests that memories for the details of experiences can be enhanced by asking a series of context-setting questions about the distinctive circumstances associated with particular episodes. Such questions might include where the respondent lived at the time, the kind of work he did, whether there was a precipitating event that brought on the episode, and the like. Experimental studies of the recall of the details of validated experiences such as doctor visits and videotaped reenactments of witnessing criminal victimizations show that these contextual questions do, in fact, help respondents recover details for episodic memories that are otherwise not remembered (Fisher & Geiselman, in press; Means & Loftus, 1991).
The situation is quite different when no single episode is distinguishable and the respondent has only a semantic memory of a prototypical episode. The best one can hope for in this case is to use the semantic memory as the basis for symptom reporting. Even then, there is good reason to believe that semantic memories may provide quite an accurate portrait of the consistent features of repeated experiences.

4.3. Remembering age-of-onset

There is some evidence in the literature to suggest that first experiences have a special place in autobiographical memory and that some of these memories are quite vivid (Pillemer, Goldsmith, Panter, & White, 1988; Robinson, 1992; Schuman, Belli, & Bischoping, under review). It is likely, though, that the extent to which this is true varies depending on a number of factors. Our analysis of the NCS question "Can you remember your exact age the very first time ...?" suggests that vivid memories of first experiences are expectedly more common for acute-onset disorders (e.g., panic and PTSD) than for disorders with insidious onsets (e.g., dysthymia, generalized anxiety disorder), and decrease with the total number of lifetime episodes of the disorder. It is unclear from the literature whether the use of pace, exhortations to think hard, or memory cues would be effective in recovering memories of first episodes for the latter disorders. If not, lower bound estimates of age-of-onset obtained by asking about clear recall of early episodes may be the best we can hope to accomplish.

Not all respondents who can recall their first episode of a disorder can remember their age when it occurred. We were not sufficiently attentive to this in the NCS pilot studies and, as a result, we combined what should have been two separate questions into a single question about clearly remembering the age of the first episode in the NCS. The two questions we should have asked are (a) "Can you remember your very first episode?" and, if so, (b) "Can you clearly remember your age at the time of this episode?"

In cases where there is a memory of an event occurring but not of when it occurred, a number of strategies are available to improve recall of the date. One involves the use of the same sort of contextual cues described above, where the interviewer asks a series of questions (e.g., "Where did you live at the time?" "Was there some crisis or stress that brought on the episode?" Etc.) in order to assist in the recall of distinctive contexts that may be associated with when the episode occurred. A danger in using this strategy alone is that a phenomenon called forward "telescoping" can lead to underestimation of how long ago an experience occurred due to the respondent inferring that an event which can be recalled vividly probably did not occur as long ago as it actually did (Brown, Rips, & Shevell, 1985). Therefore, whenever contextual cues are used to improve the vividness of recall of a first episode it is also important to use techniques that combat the potential problem of telescoping. One way this can be done is by forcing respondents to take their time and consider carefully the implications of their recollections about context for their estimation of age (Bradburn & Bradburn, 1982). If they can remember that the episode occurred shortly after they had a problem with their teacher Mrs. Smith and that Mrs. Smith was their fourth grade teacher, accuracy of dating their age at the time will be improved substantially. The use of landmark events and a time line calendar can also help reduce telescoping (Brown & Harris, 1978; Kessler & Wethington, 1991).

4.4. Remembering illness course

As noted above, research shows that when people are asked difficult recall questions they generally base their answers on estimates rather than recall. This tendency increases as the recall task becomes more difficult Bradburn and his associates (1987), for example, reported results of a study in which respondents were asked how often they ate at a restaurant in the past two months. Respondents who did not eat out very often were more likely to tally the occasions to arrive at their answer. Counting became even less common when the recall period was extended to six months.

Empirical studies show that estimation usually proceeds by the respondent combining whatever recall he may have of salient facts with semantic memories for that general class of experiences and his general knowledge to infer an answer (Bradburn et al., 1987; Brown, in press). We emphasized earlier in this chapter that this can lead to less accurate reports than those based on active memory search, but it is also important to recognize that estimation often yields fairly accurate responses to general questions that do not require recall of particular episodes (Pearson et al., 1992). In cases where episodic memories cannot be recovered even after extensive memory search -- such as when a question is asked about the number of times one ate at a restaurant in the past six months or when an adult is asked about their drinking behavior when they were in high school -- estimation may actually be more desirable (Blair & Burton, 1987, Burton & Blair, 1991). The precise inference strategies used to make the estimation are still unspecified, however. Some inference strategies may be more accurate than others and the researcher can improve accuracy of reporting by discovering the
most effective strategies and coaching the respondents to use them. However, this requires a prior program of basic research on the accuracy of alternate strategies of inferring answers to complex memory questions. No research has evaluated how respondents answer complex questions such as the following CIDI questions on the course of depression:

In your lifetime, how many spells like that (when you felt depressed and had some of these other problems like LIST SYMPTOMS OF DEPRESSION) have you had that lasted two weeks or more? Between (any of) these spells were you feeling OK at least for some months? Between (any of) these spells were you fully able to work and enjoy being with other people? Did that "normal" period last at least six months? (IF NOT) Did it last at least two months?

It is not clear how respondents answer these questions. If they only had one or two episodes in their lifetime it is likely that they will refer to episodic memories, but estimation is more likely if the number of episodes is large (Burton & Blair, 1991). Indeed, our debriefing work for the NCS shows that this is the case. The respondents who reported a small number of episodes generally said that they counted to answer the question about number of episodes, while those who reported a large number said that they estimated. Based on the work of Means and Loftus (1991), it is likely that we could improve the accuracy of reports about duration and time between episodes by respondents who referred to episodic memories by asking a series of concrete questions that help discriminate the episodes in the respondent's mind. If we had a better understanding of the accuracy of different estimation methods, we might also be able to improve the accuracy of the reports given by respondents who use estimation by decomposing the questions and guiding the estimation process.

5. Overview

Before closing, it must be noted that we explicitly ignored the difficult question of whether respondents are honest either with the interviewer or with themselves in discussing their mental health. The issue of honesty is a problematic one. The methodological literature on the accuracy of respondent reports shows clearly that the perceived social desirability of responses is as important as understanding or memory in determining the accuracy of reports (Sudman & Bradburn, 1974; Kessler & Wethington, 1991). We have no way of assessing the magnitude of this problem with available data, but it clearly needs to be taken into consideration as research in this area moves forward.

Focusing only on the ability to recall rather than on willingness to disclose memories, the chapter reviewed a large body of research suggesting that it might be possible to recover long-term memories of salient aspects of childhood psychiatric disorders with active memory search. Other memories may be too difficult to recover, either because they were never salient or because of more active processes that might involve repression. It is important to recognize, though, that very few validation studies have been carried out concerning long-term memories of any sort and none at all concerning reports about childhood psychiatric disorders. Studies of this sort must be conducted to advance our understanding beyond speculation.

Two considerations support our belief that validation studies are critical. First, it is quite likely that the use of careful question wording and field administration procedures could be shown by such studies to yield more valid retrospective data about childhood psychiatric disorders. Second, as the NCS revealed alarmingly high rates of untreated symptoms among the general population (Kessler et al., 1994), we must expedite the search for antecedents to these problems; interventions simply cannot wait for long-term prospective and experimental studies to advance our understanding of developmental psychopathology. While retrospective reports have limitations, as this review makes amply clear, such data nonetheless are an invaluable supplement to more time and cost intensive methods. The history of epidemiologic research in other substantive areas is testimony to this fact. For example, the discovery of the effects of tampon use on toxic shock syndrome and of cigarette smoking on lung cancer were both initially revealed during retrospective studies (Schlesselman, 1982). The challenge for researchers who seek to understand childhood psychiatric disorder, its causes and sequelae, is to understand the legitimate uses and limits of retrospective data.
REFERENCES


