The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI)

RONALD C. KESSLER,[1] T. BEDIRHAN ÜSTÜN[2]

[1] Department of Health Care Policy, Harvard Medical School, Boston MA, USA
[2] Global Programme on Evidence for Health Policy, World Health Organization, Geneva, Switzerland

ABSTRACT *This paper presents an overview of the World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI) and a discussion of the methodological research on which the development of the instrument was based. The WMH-CIDI includes a screening module and 40 sections that focus on diagnoses (22 sections), functioning (four sections), treatment (two sections), risk factors (four sections), socio-demographic correlates (seven sections), and methodological factors (two sections). Innovations compared to earlier versions of the CIDI include expansion of the diagnostic sections, a focus on 12-month as well as lifetime disorders in the same interview, detailed assessment of clinical severity, and inclusion of information on treatment, risk factors, and consequences. A computer-assisted version of the interview is available along with a direct data entry software system that can be used to keypunch responses to the paper-and-pencil version of the interview. Computer programs that generate diagnoses are also available based on both ICD-10 and DSM-IV criteria. Elaborate CD-ROM-based training materials are available to teach interviewers how to administer the interview as well as to teach supervisors how to monitor the quality of data collection.*

Key words: Composite International Diagnostic Interview, epidemiologic research design, psychiatric diagnostic interview, question wording methods

Introduction

This paper discusses methodological issues involved in designing the World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI), the interview used in the US National Comorbidity Survey Replication (NCS-R; Kessler and Merikangas, 2004). The WMH-CIDI was developed by WHO for use in the WHO WMH Survey Initiative. The latter is a series of face-to-face household surveys carried out with coordination by WHO in 28 countries around the world (Kessler, 1999; Kessler and Üstün, 2000). These surveys aim to obtain valid information about the prevalence and correlates of mental disorders in the general population, unmet need for treatment of

mental disorders, treatment adequacy among patients in treatment for mental disorders, and the societal burden of mental disorders. The focus of the current paper is on the sections of the WMH-CIDI that assess psychopathology, although a few words also are said about other sections of the instrument.

Historical overview

The first fully structured psychiatric diagnostic interview that could be administered by trained lay interviewers was the Diagnostic Interview Schedule (DIS) (Robins, Helzer, Croughan and Ratcliff, 1981). The DIS was developed by Lee Robins and her colleagues at Washington University with support from the National Institute of Mental Health for use in the Epidemiologic Catchment Area (ECA) Study

(Robins and Regier, 1991). The ECA was a landmark community-based survey of mental disorders carried out in selected neighbourhoods in five US communities. The wide dissemination of ECA results in high-profile publications led to replications in other countries as well as to the development of other structured diagnostic interviews. The most widely used of these instruments is the WHO CIDI (World Health Organization, 1990). The CIDI is an expansion of the DIS that was developed under the auspices of WHO by an international task force under the supervision of Lee Robins to address the problem that DIS diagnoses are exclusively based on the definitions and criteria of the American Psychiatric Association's (APA) Diagnostic and Statistical Manual (DSM) of Mental Disorders (Robins, Wing, Wittchen, Helzer, Babor, Burke, Farmer, Jablenski, Pickens, Regier, Sartorius and Towle, 1988). The WHO was keen to expand the DIS to generate diagnoses based on the definitions and criteria of the WHO International Classification of Disease (ICD). This was especially important for cross-national comparative research, as the ICD system is the international standard diagnostic system.

The CIDI was designed to encourage community epidemiological surveys in many countries around the world. To this end, a multinational CIDI editorial committee translated and field-tested the instrument in many different countries (Wittchen, 1994), while WHO encouraged researchers around the world to carry out CIDI surveys beginning in 1990 when the CIDI was first made available. These efforts were successful as over a dozen large-scale CIDI surveys in as many countries were completed during the first half of the 1990s. The WHO created the International Consortium in Psychiatric Epidemiology (ICPE) in 1997 to bring together and compare results across these surveys (Kessler, 1999). The ICPE has subsequently published a number of useful descriptive studies of cross-national similarities and differences in prevalence and socio-demographic correlates of mental disorders (for example, Aguilar-Gaxiola, Alegria, Andrade, Bijl, Caraveo-Anduaga, DeWit, Kolody, Kessler, Üstün, Vega and Wittchen, 2000; Alegria, Kessler, Bijl, Lin, Heeringa, Takeuchi and Kolody, 2000; Bijl, de Graaf, Hiripi, Kessler, Kohn, Offord, Üstün, Vicente, Vollebergh, Walters and Wittchen, 2003; WHO International Consortium in Psychiatric Epidemiology, 2000).

However, the work of the ICPE with this first generation of CIDI surveys was hampered by the fact that comparability among the surveys was limited to the assessment of mental disorders. Measures of risk factors, consequences, patterns and correlates of treatment, and treatment adequacy, none of which were included in the CIDI, were not assessed in a consistent manner across the surveys.

Recognizing the value of coordinating the measurement of these broader areas of assessment, the ICPE launched an initiative in 1997 to bring together the senior scientists in planned CIDI surveys prior to the time their surveys were carried out in order to coordinate measurement. Within a short period of time, research groups in over a dozen countries joined this initiative. The World Health Organization officially established the WHO WMH Survey Initiative to coordinate this undertaking in 1998. Since that time, the number of participating WMH countries has expanded to 28 with an anticipated combined sample size of over 200,000 interviews. The authors of the current paper are the co-directors of both the ICPE and the WMH Survey Initiative as well as the principal developers of the WMH-CIDI, the expanded version of the WHO CIDI that was created for use in the WMH surveys.

An overview of the WMH-CIDI

In the course of expanding the CIDI to include broader areas of assessment, we also took the opportunity to make the diagnostic sections of the CIDI more operational. We expanded questions to break down critical criteria, including the clinical significance criteria required in the DSM-IV system. We expanded the diagnostic sections to include dimensional information along with the categorical information that existed in previous CIDI versions. We also expanded the number of disorders included in the CIDI.

The 41 sections in the WMH-CIDI are listed in Table 1. These are not in their order of assessment. The first section is an introductory screening and lifetime review section, the logic of which is discussed later in this article. There are also 22 diagnostic sections that assess mood disorders (two sections), anxiety disorders (seven sections), substance-use disorders (two sections), childhood disorders (four sections), and other disorders (seven sections). Four additional sections assess various

Table 1. An outline of the WMH-CIDI

I. Screening and lifetime review	
II. Disorders	
Mood	Major Depression, Mania
Anxiety	Panic Disorder, Specific Phobia, Agoraphobia, Generalized Anxiety Disorder, Post-Traumatic Stress Disorder, Obsessive-Compulsive Disorder, Social Phobia
Substance abuse	Alcohol Abuse, Alcohol Dependence, Drug Abuse, Drug Dependence, Nicotine Dependence
Childhood	Attention-Deficit/Hyperactivity Disorder, Oppositional-Defiant Disorder, Conduct Disorder, Separation Anxiety Disorder
Other	Intermittent Explosive Disorder, Eating Disorders, Premenstrual Disorder, Non-Affective Psychoses Screen, Pathological Gambling, Neurasthenia, Personality Disorders Screens
III. Functioning and physical disorders	Suicidality, 30-day Functioning, 30-day Psychological Distress, Physical Comorbidity
IV. Treatment	Services, Pharmacoepidemiology
V. Risk factors	Personality, Social networks, Childhood experiences, Family Burden
VII. Socio-demographics	Employment, Finances, Marriage, Children, Childhood Demographics, Adult Demographics
VII. Methodological	Part I – Part II Selection, Interviewer Observations

kinds of functioning and physical comorbidity. Two assess treatment of mental disorders. Four assess risk factors. Six assess socio-demographics. Two final sections are methodological. The first of these two includes rules for determining which respondents to select into Part II of the interview and which ones to terminate after Part I of the interview. The second methodological section consists of interviewer observations that are recorded after the interview has ended.

The entire WMH-CIDI takes an average of approximately 2 hours to administer in most general population samples. However, interview time varies widely depending on the number of diagnostic sections for which the respondent screens positive. As mentioned in the last paragraph, the interview has a two-part structure that allows early termination of a representative subsample of respondents who show no evidence of lifetime psychopathology. The sampling fraction used in this subselection procedure influences average interview time. Finally, a number of WMH-CIDI sections are optional and can be administered to subsamples rather than to the entire sample. This, too, reduces average interview length.

In addition to the interview schedule, we developed an elaborate set of training materials to teach interviewers how to administer the WMH-CIDI and to teach supervisors how to monitor the quality of data collection. We developed a computer-assisted version of the interview (CAI) that can be used with laptop computers. We also developed a direct data entry (DDE) software system that can be used to keypunch paper and pencil versions of the interview. Finally, we developed computer programs that generate diagnoses from the completed survey data using the definitions and criteria of the ICD-10 or the DSM-IV diagnostic systems.

Use of the WMH-CIDI requires successful completion of a training programme offered by an official WHO CIDI Training and Research Centre (CIDI-TRC). Another innovation associated with the WMH-CIDI is a state-of-the art interviewer

training programme that includes an intelligent 40-hour CD-ROM-based self-study module in addition to a three-day face-to-face training module that requires the trainee to travel to an authorized CIDI-TRC. The latter is designed for individuals who have successfully completed the self-study module, as indicated by passing the self-administered tests embedded throughout the CD-ROM. Remedial training elements are embedded in the CD-ROM whenever a trainee fails an embedded test. Trainees who successfully complete the certification process at the end of this program are given access to all WMH-CIDI training materials for use in training interviewers and supervisors. They are also given copies of the WMH-CIDI CAI and DDE programs and the computerized diagnostic algorithms. A PDF copy of the WMH-CIDI and contact information for WMH-CIDI training can be obtained from the CIDI Web page at www.who.int/msa/cidi.

The validity of CIDI diagnostic assessments
A number of DIS and CIDI validity studies were carried out prior to the time the WMH Survey Initiative was launched. These studies aimed to determine whether the diagnoses generated by these instruments are consistent with those obtained independently by trained clinical interviewers who administer semi-structured research diagnostic interviews to a probability sample of survey respondents who previously completed the DIS or CIDI. Wittchen (1994) reviewed these studies up through the early 1990s. Only a handful of DIS or CIDI validity studies have been published since that time (Kessler, Wittchen, Abelson, McGonagle, Schwarz, Kendler, Knauper and Zhao, 1998; Wittchen, Üstün and Kessler, 1999; Brugha, Jenkins, Taub, Meltzer and Bebbington, 2001). Results show that DIS and CIDI diagnoses are significantly related to independent clinical diagnoses, but that individual-level concordance is far from perfect. Some part of this lack of concordance is doubtless due to unreliability of clinical interviews. Indeed, the literature is clear in showing that test-retest reliability is higher for diagnostic classifications based on DIS-CIDI interviews than semi-structured clinical interviews. However, there is also the issue of validity, which is presumably higher in semi-structured clinical interviews than in fully structured DIS-CIDI interviews. As a result of concerns about validity, considerable

interest existed among the developers of the WMH-CIDI to improve the validity of the CIDI for use in the WMH surveys.

Based on previous evaluations of the CIDI by survey methodologists in preparation for the US National Comorbidity Survey (NCS) (Kessler, Wittchen et al., 1998; Kessler, Mroczek and Belli, 1999; Kessler, Wittchen, Abelson and Zhao, 2000), four main methodological problems were the focus of our work revising the diagnostic sections of the WMH-CIDI. One was that respondents might not understand some of the CIDI questions, a number of which included multiple clauses and vaguely defined terms. A second was that some respondents might not understand the task implied by the questions, which sometimes required careful memory search that was unlikely to be carried out unless respondents were clearly instructed to do so. A third was that respondents might not be motivated to answer accurately, especially in light of the fact that many CIDI questions deal with potentially embarrassing and stigmatizing experiences. A fourth was that respondents might not be able to answer some CIDI questions accurately, especially those that asked about characteristics of mental disorders that are difficult to remember (for example, age of onset, number of lifetime episodes).

A considerable amount of methodological research has been carried out by survey researchers on each of the four methodological problems enumerated in the last paragraph (for example, Turner and Martin, 1985; Tanur, 1992; Sudman, Bradburn and Schwarz, 1996). This research has advanced considerably over the past two decades as cognitive psychologists have become interested in the survey interview as a natural laboratory for studying cognitive processes (Schwarz and Sudman, 1994; 1996; Sirken, Herrmann, Schechter, Schwarz, Tanur and Tourangeau, 1999). A number of important insights have emerged from this work that suggest practical ways of improving the accuracy of self-reported psychiatric assessments. As described below, we used these insights to help develop the WMH-CIDI. The next four sections of the paper provide a quick review of these insights as well as use of them to address each of the four methodological problems enumerated in the last paragraph.

Question comprehension

It is obvious that ambiguous questions are likely to be misconstrued. It is perhaps less obvious, though, just how ambiguous most structured questions are and how often respondents must 'read between the lines'. In the first systematic study of this issue, Belson (1981) debriefed a sample of survey respondents on a set of standard survey questions and found that more than 70% of respondents interpreted some questions differently from the researcher, leading Belson to conclude that subtle misinterpretations are pervasive in survey situations. Similar conclusions have been obtained in other survey debriefing studies (Oksenberg, Cannell and Kanton, 1991). Our own debriefing studies of the CIDI found much the same result – a great many respondents misunderstood important aspects of key diagnostic questions.

How is it possible for so much misunderstanding to occur? As Oksenberg and her colleagues discovered, the answer lies partly in the fact that many terms in surveys are vaguely defined. Beyond this is the more fundamental fact that the survey interview situation is a special kind of interaction in which the standard rules of conversation – rules that 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 the person to whom he or she is talking or the purpose of the conversation (Cannell, Fowler and 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 survey 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 leads to more misreading than in normal conversations even when questions are seemingly straightforward (Clark and Schober, 1992), a problem that is compounded when the topic of the interview is one that involves emotional experiences that are in many cases difficult to describe with clarity.

Clinical interviews attempt to deal with this problem by being 'interviewer based' (Brown, 1989); that is, by training the interviewer to have a deep understanding of the criteria being evaluated, allowing the interviewer to query the respondent as much as necessary to clarify the meaning of questions, and leaving the ultimate judgment about the rating with the interviewer rather than the respondent. Indeed, one might say that the interview is, in some sense, administered to the interviewer rather than to the respondent in that the responses of interest are responses to interviewer-based questions of the following sort: 'Interviewer, based on your conversation with the respondent, would you say that he or she definitely, probably, possibly, probably not, or definitely does not meet the requirements of Criterion A?' Fully structured psychiatric interviews like the CIDI cannot use this interviewer-based approach because, by definition, they are designed so that interviewer judgment plays no part in the responses. These 'respondent-based' interviews use totally structured questions that the respondent answers, often in a yes-no format, either after reading the questions to themselves or after having an interviewer read the questions aloud. When the criterion of interest is fairly clear, there may be little difference between interviewer-based and respondent-based interviewing. It is a good deal more difficult, though, to assess conceptually complex criteria with fully structured questions.

In an effort to investigate the problem of question misunderstanding in the CIDI as part of the pilot studies for an early CIDI survey, Kessler and his colleagues (Kessler, Wittchen et al., 1998; Kessler et al., 1999) carried out a series of debriefing interviews with community respondents who were administered sections of the CIDI and then asked to explain what they thought the questions meant and why they answered the way they did. A great deal of misunderstanding was found. However, enormous variation across questions was also found in the frequency of misunderstanding. Four discriminating features were found among questions that had high versus low levels of misunderstanding.

First, some commonly misunderstood questions are simply too complex for many respondents to grasp. Second, some commonly misunderstood questions involve vaguely defined terms rather than complex concepts. A third type of commonly misunderstood CIDI question involves questions about odd experiences that could plausibly be interpreted in more than one way, such as being asked about seeing

and hearing things that others do not. Many respondents have a tendency to normalize these questions and respond positively when the correct answer is negative. A fourth type of commonly misunderstood CIDI question, finally, involves a contextual misunderstanding – that is, a misunderstanding that derives more from the position of the question in the flow of the interview than from lack of clarity in the question. A good example is the evaluation of Criterion A in the DSM-III-R diagnosis of simple phobia, which stipulates that the fear of circumscribed stimuli must be 'persistent'. CIDI 1.0 operationalized this criterion by asking 'Did this strong unreasonable fear continue for months or even years?' Although seemingly not ambiguous in itself, pilot work by Kessler and his colleagues found that this question was misunderstood by a great many respondents because of the location of the question in the instrument. Specifically, this question followed an open-ended question that asked the respondent to give an example of a specific fear. In many cases the respondent would respond to this open-ended question by describing the autonomic arousal symptoms that occur on exposure to the stimulus such as feeling dizzy or having trouble breathing. When the follow-up question was administered right after this description – asking whether this fear continued for months or even years – the question was sometimes misunderstood as asking about the duration of the arousal symptoms. The respondent would invariably answer no, the nausea or dizziness or other physiological symptoms typically lasted no more than a few hours and certainly never lasted months or years. Confusions of this sort can lead to substantial errors in fully structured clinical assessments, especially when they concern required symptoms.

Following the work of Kessler et al. (2000), we carried out detailed CIDI debriefing interviews with volunteer respondents in methodological studies of the WMH-CIDI in an effort to pinpoint CIDI questions with each of the above four types of comprehension problems. Misunderstandings based on complex questions are addressed by breaking down the original CIDI questions into less complex subquestions. Especially complex questions are presented in a respondent booklet (RB) that provides a visual aid as the questions are being read by the interviewer. Misunderstandings based on the vagueness of terms are addressed by introducing clar-

ifications and examples. Misunderstandings based on normalization of questions about odd experiences are addressed by prefacing the questions with clarification that we are actually asking about odd experiences and that it is important for us to learn how often these experiences occur. Contextual misunderstandings, finally, are resolved by reordering questions to remove the contextual effects and by adding clarifying clauses in questions where residual confusion might exist. Although the number of modifications of this sort are so large that each one cannot be reviewed here, the appendix to this paper presents one example of each of the four types of modifications to illustrate the types of changes we made to the CIDI in developing the WMH-CIDI.

Task comprehension

Respondents not only sometimes misunderstand survey questions, but they also sometimes misunderstand the fundamental task they are being asked to carry out. Debriefing studies have shown that misunderstandings of the second sort are especially common with the diagnostic stem questions in the CIDI (Kessler et al., 2000). These stem questions are the first questions asked in each diagnostic section. They are used to determine whether a lifetime syndrome of a particular sort might have ever occurred. The questions provide what are, in effect, brief vignettes and ask the respondent whether they ever had an experience of this sort. If so, additional questions assess the specifics of the syndrome. If not, the remaining questions about this syndrome are skipped. Substantial confusion arises from respondents' failure to understand the purpose of such stem questions. In particular, only about half of pilot respondents in the Kessler study interpreted these questions as they were 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 early research on standard interview practice, respondents are generally ill-informed about the purposes of the research and poorly motivated to participate actively. Furthermore, cues from interviewers often reinforce this inclination to participate in a half-hearted way. For example, 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. 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 instruction can substantially improve the quality of data obtained in surveys. Based on this result, we built in clarifying statements throughout the WMH-CIDI aimed at informing respondents that accuracy was important. See the appendix for an example of such a statement.

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 methodologists has led to the development of motivational techniques 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.

Motivational instructions

There is evidence that the use of introductory remarks at the beginning of a survey that clarify the research aims can help motivate respondents to provide a more complete and accurate report than they would otherwise (Cannell et al., 1981). Debriefing shows that respondents are more willing to undertake laborious and possibly 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.' Based on this evidence, we developed and presented a statement containing a clear rationale for administering the interview at the onset of the WMH-CIDI interview schedule and emphasized the importance of the survey for social policy purposes. See the appendix for this text. In addition, in the case of especially important questions that require long-term recall to answer correctly, the questions are included in the RB and a written instruction is included at the top of the page in capital letters urging the respondent to 'take your time and think carefully before answering'.

Contingent reinforcement

Consistent with research on behavioural modification of verbal productions through reinforcement (see, for example, Centers, 1964), several survey researchers have demonstrated that verbal reinforcers such as 'thanks' and 'that's useful' can significantly affect the behaviour of survey respondents (Marquis et al., 1969). Based on this observation, 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 and 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 contingent feedback instruction, interviewers issue some expression of gratitude whenever the respondent seems to consider his or her answer carefully, whether they remember 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 gives an immediate answer whether or not anything was reported.

Experiments carried out by Cannell and his associates (Miller and Cannell, 1977; Vinokur, Oksenberg and 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 behaviours (for example, gynaecological problems, seeing an X-rated movie) and a decreased tendency to over-report self-enhancing behaviours (for example, the number of books read in the last 3 months, reading the editorial page of the newspaper the previous day). Based on these results, the Cannell contingent feedback approach is included as a fundamental part of the interviewer training materials developed for the WMH surveys.

Commitment questions
Instructions that define the nature of interviewer expectations for respondent behaviour 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 behaviours (see, for example, Loftus and Palmer, 1974). This same literature shows that perspective can have motivational force when it implies a common purpose (Clark et al., 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 demur explicitly or tacitly 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 honour the injunction implied in the perspective of the question and this implied commitment, in turn, creates motivation to this task (Marlatt, 1972). Based on this type of thinking, Cannell and his colleagues showed that it is possible to motivate respondents to accept the goal of serious and active reporting by asking an explicit commitment question as part of the inter-

view. Experimental studies carried out by Cannell and his associates (Oksenberg et al., 1979a; Oksenberg, Vinokur and Cannell, 1979b; Cannell et al., 1981) have shown that commitment questions improve accuracy of recall. Based on this result, we added a commitment question in the screening section of the WMH-CIDI just before the administration of lifetime diagnostic stem questions. See the appendix for the text of this question.

The ability to answer accurately

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 and Zajonc, 1985). When the respondent has a history of many instances of the same experience that 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 and Loftus, 1990; Means and 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 probably 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 and Steele, 1984; Skelton and 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 (Brewer, 1986; Belli, 1988; Menon, 1994).

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 is being asked to recover an episodic memory. Furthermore, 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 and 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. Therefore, as part of the WMH-CIDI pilot work, we debriefed pilot respondents with an explicit eye towards pinpointing questions that were difficult to answer. When questions of this sort were discovered, an attempt was made to revise the questions to reduce the memory problem either by allowing explicitly for estimation (for example, explicitly asking respondents to provide a rough estimate), by providing categorical responses that reduce the complexity of the task, or by decomposing the question into sub-questions that mimic effective memory search processes. Examples of question wording modifications that employ each of these strategies are presented in the appendix.

The lifetime review section
The previous four sections of the paper review a number of strategies that we use to optimize data quality in the WMH-CIDI either by improving understanding, by enhancing respondent commitment, or by adjusting questions to recognize that some respondents will be less able than others to provide completely accurate responses. We use some additional strategies to deal with two or more of these problems at once. The most important of the latter was a life review section that we administer near the beginning of the interview in an effort to both motivate and facilitate active memory search in answering diagnostic stem questions. This section starts out with an introduction that explains to respondents that the questions might be difficult to answer because they require respondents to review their entire lives. The introduction then goes on to say that despite this difficulty it was very important for the research that these questions be answered accurately. The introduction ends with the injunction to 'please take your time and think carefully before answering' and a commitment question that asks respondents if they were willing to do this.

The diagnostic stem questions for all core diagnoses are administered directly following the commitment question. The questions are all included in the RB with a written instruction to 'take your time and think carefully before answering'. Interviewers were instructed to read the diagnostic stem questions slowly in an effort to emphasize their importance and to use motivational probes to encourage active memory search. Our intent in developing this section was that we could both explain the serious and difficult nature of the task and motivate respondents to engage in active memory search we hoped to stimulate by combining all the stem questions after a fairly detailed motivational introduction. We also recognize, based on our debriefing studies, that CIDI respondents quickly learn the logic of the stem-branch structure after a few sections and recognize that they can shorten the interview considerably by saying no to the stem questions. This problem has been removed by asking all the stem questions near the beginning of the interview before the logic of the stem-branch structure became clear. Furthermore, respondents told us in debriefing interviews that their energy flagged as the interview progressed, making it much more difficult to carry out a serious memory search later in the interview than at the beginning.

As previously noted, Cannell and his associates carried out experiments that documented powerful

effects of commitment questions on the accuracy of survey responses (Oksenberg et al., 1979a; 1979b; Cannell et al., 1981). A similar experiment was carried out in conjunction with the NCS to evaluate the effects of using a lifetime review section in conjunction with commitment and motivational probes. A random sample of 200 community respondents was randomized either to the standard version of the CIDI or to a version that was identical except that it included the life review section. As reported in more detail elsewhere (Kessler, Wittchen et al., 1998), this experiment documented that the life review section led to a significant increase in the proportion of respondents who endorsed diagnostic stem questions. For example, while 26.7% of respondents in the standard CIDI condition endorsed the 'sad, blue, or depressed' stem question for major depression, a significantly higher 40.6% did so in the life-review condition. Importantly, a clinical validity study documented that this increased prevalence of stem endorsement was not accompanied by a reduction in sensitivity with regard to clinical diagnoses, documenting that additional true cases were discovered by the use of the life review section and the accompanying commitment and motivation probes. These results were the basis of adopting the lifetime review section in the WMH-CIDI.

Substantive modifications of diagnostic sections
In addition to the methodological modifications described above, a number of important substantive modifications were made to the diagnostic assessments in the WMH-CIDI aimed at addressing current uncertainties about the prevalence, impairment, and appropriate diagnostic criteria for the disorders assessed in the interview. Perhaps the most important of these uncertainties concerns diagnostic thresholds. This uncertainty arose, in no small part, as a reaction to the results of early DIS and CIDI surveys, which showed that as much as 50% of the general population of some countries meet lifetime criteria for one or more ICD or DSM mental disorders and as many as 30% meet criteria for such a disorder in the past 12 months (WHO International Consortium in Psychiatric Epidemiology, 2000; Bijl et al., 2003). These percentages seemed implausibly high to many critics, leading to the suggestion that the lay-administered diagnostic interviews in these surveys were upwardly biased (Brugha, Bebbington

and Jenkins, 1999; Wittchen et al., 1999). However, clinical calibration studies showed that the prevalence estimates in these surveys were not upwardly biased (Kessler, Wittchen et al., 1998; Eaton, Neufeld, Chen and Cai, 2000), leading critics to conclude that the ICD and DSM systems themselves are overly inclusive (Pincus, Zarin and First, 1998; Regier, Kaelber, Rae, Farmer, Knauper, Kessler and Norquist, 1998; Üstün, Chatterji and Rehm, 1998).

This conclusion was instrumental in causing an APA task force to add a clinical significance criterion to many disorders in the DSM-IV in order to remind readers of the basic definition of a mental disorder in the introduction of the manual as requiring clinically significant distress or impairment. However, even when this additional requirement was applied post hoc to DIS and CIDI surveys carried out in the US, the 12-month prevalence of having at least one DSM disorder, equivalent to approximately 37 million adults in the US, continued to substantially exceed the number who could be helped with current treatment resources (Narrow, Rae, Robins and Regier, 2002). In recognition of this problem, several more restrictive definitions have been proposed that can be used to narrow the number of people qualifying for treatment (National Advisory Mental Health Council, 1993; Substance Abuse and Mental Health Services Administration, 1993; Regier, 2000; Narrow et al., 2002; Regier and Narrow, 2002).

Others, however, have argued against the proposal to narrow the definition of mental disorders (Mechanic, 2003) and have, in some cases, even argued that the definitions should be expanded to include what would currently be considered subthreshold cases (see, for example, Merikangas, Zhang, Avenevoli, Acharyya, Neuenschwander and Angst, 2003). These critics have noted that research shows many syndromes currently defined as mental disorders to be extremes on continua that appear not to have meaningful thresholds (see Preisig, Merikangas and Angst, 2001). This is important for research purposes because exploration of the full continua rather than the currently established diagnostic thresholds yields greater power in studies of genetic and environmental risk factors (Benjamin, Ebstein and Lesch, 1998). With regard to diagnostic thresholds, these critics note that research has shown subthreshold cases on some of these continua to be

quite impaired (for examples, Judd, Paulus, Wells and Rapaport, 1996) and to have significantly elevated risk of serious outcomes such as suicide attempts and hospitalization for emotional problems (Kessler, Barber, Beck, Berglund, Cleary, McKenas, Pronk, Simon, Stang, Üstün and Wang, 2003). This means that the development of early interventions to treat these subthreshold cases might prevent progression along a given severity continuum, thereby reducing the prevalence of serious cases in a cost-effective fashion (Eaton, Badawi and Melton, 1995). Removal of these subthreshold cases from the ICD or DSM systems, in comparison, might result in the importance of developing interventions for these cases to be ignored as well as to a distortion occurring in the reality that mental disorders, like physical disorders, vary widely in seriousness (Spitzer, 1998; Kendell, 2002).

The final adjudication between these competing views will doubtless take years to occur and will rely on emerging information about the genetics of mental disorders as well as on information about treatment response across the range of the symptom severity continuum. To the extent that epidemiological data will play a part, they will, at a minimum, need to include assessments of subthreshold cases, assessments of symptom severity in dimensional terms, and evaluation of the association between symptom severity and impairment. The WMH-CIDI is designed to do all three of these things, as briefly discussed in the next three subsections.

Subthreshold disorders
Our general approach in modifying WMH-CIDI diagnostic sections is to include as much information as possible about subthreshold cases, with the precise nature of the subthreshold assessment guided by the literature and our preliminary studies. For example, in the case of depression, even though the diagnostic criteria require dysphoria or anhedonia that persists most of the day, we found that a great many people otherwise meet the criteria for a major depressive episode except that their symptoms persist only for about half the day or sometimes less than half the day. We include these people in our assessment of depressive disorders. In addition, we include people with as few as two symptoms in their worst lifetime episode of depression in order to capture cases of minor depression even though larger numbers are

required in the ICD and DSM systems to be considered a major depressive episode. We also include people with depressive episodes as short as 3 days if they report ever having episodes of this sort most months for an entire year in a row. This is done to allow an assessment of recurrent brief depression (Angst, Merikangas, Scheidegger and Wicki, 1990).

To take a second example, in the case of panic we carry out a complete assessment of people who report at least one lifetime limited symptom attack or panic attack. Age of onset and the circumstances surrounding the single attack were collected. Parallel information is collected about the first lifetime attack among people who meet criteria for a panic disorder. Separate information is obtained from the latter people about age of onset of the transition from panic attacks to panic disorder. Previous research has shown that only about half of the people who have a single lifetime panic attack go on to develop panic disorder (Eaton, Kessler, Wittchen and Magee, 1994). The information collected in the WMH-CIDI allows us to obtain separate prevalence estimates of panic attacks and panic disorder as well as to study the separate predictors of progression from panic attacks to panic disorder. A similar distinction is made between specific fears and phobias, with separate dating of age of onset of the fear, and avoidance.

As a final example, we include a complete assessment of generalized anxiety disorder (GAD) for respondents who report having episodes that persist at least one month, the original duration requirement in DSM-III, rather than requiring the 6-month minimum episode duration stipulated in DSM-IV and ICD-10. The decision by the developers of the DSM to increase the 1-month minimum duration requirement for GAD in DSM-III to six months in DSM-III-R was based on the fact that the vast majority of patients with GAD in treatment samples had comorbid depression unless their episodes of GAD persisted for at least six months (Breslau and Davis, 1985). However, subsequent epidemiological research showed that pure cases of GAD with shorter durations exist in the general population, but seldom come to clinical attention because professional help-seeking is often driven by comorbid disorders. Yet this does not mean that people with recurrent episodes of GAD lasting less than 6 months are not impaired. Indeed, the largest and most comprehensive study of this matter, carried out by Maier et al.

(2000) in the WHO study of mental disorders in primary care, found that the 1-month duration requirement is optimal for distinguishing cases of GAD from non-cases in terms of role impairment. The decision to require a minimum duration of 1 month rather than 6 months in the WMH-CIDI is based on this result. For a similar reason we assess cases of subthreshold GAD who fail to report all the psycho-physiological symptoms required by DSM-IV or ICD-10.

Symptom persistence and severity

The focus of the CIDI, like that of the DIS before it, has largely been on lifetime disorders, although a 12-month version of the CIDI was developed in the last revision of the instrument. The standard lifetime version of the CIDI provides only superficial information about recent disorders by asking no more than a single question – 'How recently have you had [the disorder]?' – to learn about recency after the assessment of lifetime symptom clustering. This makes it impossible to characterize the persistence of disorders over the recent past or to know whether respondents with a lifetime disorder have met full criteria during the recent past. As 12-month prevalence is of great interest for needs assessment, this superficial consideration of 12-month prevalence is a serious limitation. The WMH-CIDI addresses this problem by obtaining information about 12-month symptoms and persistence of symptoms over the past year. In the case of panic and intermittent explosive disorder, this is done by asking about number of attacks in the past 12 months as well as about the number of months when the respondent had at least one attack. In the case of episodic disorders, such as depression and GAD, 12-month duration is assessed by asking how many weeks out of the past 52 the respondent has been in an episode.

Similar data are obtained in the WMH-CIDI to increase understanding of long-term course. In the standard version of the CIDI, information on course is limited to two questions about age of onset and age of recency of the disorder. The WMH-CIDI expands this assessment to ask about persistence in the interval between these two ages along the same lines used to assess 12-month persistence (for example, with questions about lifetime number of panic and anger attacks, lifetime number of episodes of depression and mania and GAD, typical and longest durations of episodes, and number of years in which the respondent experienced at least one attack or one episode of the disorder). In keeping with the prior comments on the limits of autobiographical memory, and consistent with the results of our methodological pilot studies, we recognize that respondents with complex histories of psychopathology will be unable to recover episodic memories in answering these long-term recall questions. As a result, the questions are worded in such a way as to make it clear that we are looking for semantic memories. Even with this limitation in mind, though, these data can be extremely useful in distinguishing between broad categories of people who have had only one or two, a few, more than a few, or a large number of attacks or episodes of episodic disorders.

The WMH-CIDI also includes much more extensive information on symptom severity than the standard CIDI. Each diagnostic section contains explicit questions about the depth of the distress caused by the disorder along with a 12-month symptom severity scale based on a fully structured version of a standard clinical scale. For example, the quick self-report version of the Inventory of Depressive Symptomatology (Rush, Gullion, Basco, Jarrett and Trivedi, 1996) is used to assess the severity of 12-month depression, while a structured version of the Panic Disorder Severity Scale (Shear, Brown, Barlow, Money, Sholomskas, Woods, Gorman and Papp, 1997) is used to assess the severity of 12-month panic. These scales are administered only to respondents who meet the subthreshold diagnostic requirements for the disorder in question and who report that they had symptoms in the past 12 months. Our hope in embedding these standard clinical symptom severity scales in the WMH-CIDI is that they will help create a crosswalk between the findings in epidemiological surveys and the findings in clinical studies.

Internal impairment

The issue of impairment is related to the issue of clinical significance. The standard CIDI asks only one dichotomous disorder-specific role impairment question for all disorders: 'Did (the disorder) ever interfere a lot with your life or activities?' No questions about impairment are asked independent of disorders. This is inadequate for evaluating whether

there is clinically significant role impairment associated with a particular syndrome or for investigating the implications of changing diagnostic thresholds on evaluations of impairment. As a result, the WMH-CIDI has substantially expanded the number of the within-section questions about the impairments caused by individual disorders. These are called 'internal' impairment questions because they ask respondents to evaluate the impairment caused by a given disorder. As described below, the WMH-CIDI also includes two important sections that assess 'external' impairment, by which we mean overall impairment in various areas of functioning without reference to the cause of the impairment. Although most of the WMH-CIDI internal impairment questions focus on the worst lifetime impairment due to a particular disorder, we also include five questions in each diagnostic section that assess impairment among 12-month cases. Four of these are the Sheehan Disability Scales (Leon, Olfson, Portera, Farber and Sheehan, 1997), which ask respondents to rate the impairments caused by a focal disorder during the one month in the past year when it was most severe in each of four areas of life (household duties, employment, social life, and close personal relationships) on a 0–10 scale that uses a visual analogue scale with impairment categories of none (0), mild (1–3), moderate (4–6), severe (7–9), and very severe (10). The fifth question asks respondents to estimate the total number of days out of 365 in the past 12 months when they were totally unable to work or carry out their other usual activities because of the focal disorder.

External impairment
Two sections of the WMH-CIDI assess external impairment. The first is the section on 30-day functioning, which is made up of the WHO Disability Assessment Schedule (WHO-DAS; World Health Organization, 1998; Rehm, Üstün, Saxena, Nelson, Chatterji, Ivis and Adlaf, 1999). The WHO-DAS assesses both the persistence (number of days in the past 30) and severity (during the days when difficulties in functioning occurred) of difficulties in the respondent's functioning during the 30 days before the interview due to all physical and mental health problems. The dimensions of functioning assessed in the WHO-DAS are keyed to the major categories in the WHO International Classification of Functioning, Disability, and Health (World Health Organization, 2001). The second section of the WMH-CIDI that assesses external impairment is the section on employment, which includes the WHO Health and Work Performance Questionnaire (HPQ) (Kessler, Barber et al., 2003). The HPQ is an expansion of the work impairment section of the WHO-DAS that assesses the workplace costs of illness in terms of absenteeism, decrements in performance while on the job, and critical workplace incidents (such as work-related accidents). The HPQ was developed in order to provide data to employers and government health policy makers about the indirect costs of illness on the productive capacity of the labour force.

Why assess both internal and external impairment?
It is important to obtain both internal (disorder-specific) and external (global) assessments of impairment. Disorder-specific assessments are important because they can be used to make direct comparisons among different mental and physical disorders. These direct comparisons are becoming increasingly central to healthcare resource allocation decisions as evidence-based medicine becomes the basis for more and more triage decisions. However, disorder-specific assessments are limited by the fact that they require respondents to make inferences about the cause of their impairments. This can be difficult, especially among the large number of people with comorbid conditions who might have a hard time sorting out which of their conditions causes various aspects of impairment.

It is important to obtain external assessments because they allow the researcher to overcome the limitation of disorder-specific assessments by empirically estimating the relative effects of different disorders from prediction equations in which measures of the prevalence of these disorders and their comorbidities are included as predictors of global impairment. However, as it is not possible to make detailed assessments of all possible disorders for inclusion in such prediction equations, estimates of the impairments due to specific disorders based on analysis of such equations are necessarily imperfect. Furthermore, replication of results involving the estimated effects of a focal disorder on impairment using external comparisons requires measurement of exactly the same set of control conditions across

studies. This is infeasible. As a result, the internal assessment of impairment is more feasible despite its conceptual limitations in comparison to the external assessment of impairment.

Ranking the impairments of mental and physical disorders
In order to provide comparative information on the impairments of mental and physical disorders, a checklist of chronic physical disorders is included in the WMH-CIDI. Internal impairment is assessed with the five questions described above for one randomly selected chronic condition per respondent. The random sampling strategy was used because comprehensive assessment of internal impairment for all possible chronic physical disorders would be too time-consuming for a one-session survey devoted to mental disorders. However, by taking care to carry out a random selection for each respondent from among all the conditions reported by that respondent, it is possible to weight the internal impairment data by the number of conditions reported to recover an equal-probability sample for each chronic condition for purposes of comparative assessment of within-disorder role impairments.

The chronic conditions checklist was modified from the list used in the National Health Interview Survey (NHIS) (National Center for Health Statistics, 2003) to ask about the lifetime occurrence, age of onset, and recency of commonly occurring chronic conditions that are thought to be associated with substantial role impairment. A number of methodological studies have found that such checklists yield valid data about disorders brought to medical attention or that significantly limit activities when compared to independent medical records (Halabi, Zurayk, Awaida, Darwish and Saab, 1992; Heliovaara, Aromaa, Klaukka, Knekt, Joukamaa and Impivaara, 1993; Edwards, Winn, Kurlantzick, Sheridan, Berk, Retchin and Collins, 1994; Gross, Bentur, Elhayany, Sherf and Epstein, 1996; Kriegsman, Penninx, van Eijk, Boeke and Deeg, 1996; Mackenbach, Looman and van der Meer, 1996). For example, moderate to high agreement (Cohen's κ; Cohen, 1960) has been found between self-reports and medical records regarding arthritis (κ= 0.41), asthma (κ= 0.55), diabetes (κ= 0.82), and high blood pressure (κ= 0.73) (Edwards et al., 1994). These are lower bound estimates because the medical record is not a 'gold standard', especially

for chronic conditions often not brought to medical attention (such as arthritis), for poorly defined conditions (such as back pain), and for symptom-based conditions in which the medical record merely reproduces symptoms that are based on self-report (such as chronic headaches).

In the case of symptom-based conditions, a number of more extensive scales are used instead of the single yes-no questions in the chronic conditions checklist. For example, we include a brief screening scale to assess migraines that reproduces physician diagnoses much more accurately than a single checklist question (Lipton, Dodick, Sadovsky, Kolodner, Endicott, Hettiarachchi and Harrison, 2003). Other symptom-based conditions that are assessed with screening scales include chronic fatigue syndrome, irritable bowel syndrome, insomnia, and unexplained chronic pain disorder.

Comparative analyses of internal impairment could add important information to the growing body of data that physical disorders often cause substantial role impairments (Zeiss and Lewinsohn, 1988; Wells et al., 1989; Wells, Stewart, Hays, Burnam, Rogers, Daniels, Berry, Greenfield and Ware, 1989; Stewart, Greenfield and Hays, 1989; Ormel, Von Korff, Üstün, Pini, Korten and Oldehinkel, 1994; Hays, Wells, Sherbourne, Rogers and Spritzer, 1995; van den Bos, 1995; Verbrugge and Patrick, 1995; Penninx, Beekman, Ormel, Kriegsman, Boeke, van Eijk and Deeg, 1996; Kempen, Ormel, Brilman and Relyveld, 1997; Ormel, Kempen, Deeg, Brilman, van Sonderen and Relyveld, 1998; Kempen, Sanderman, Miedema, Mayboom-de and Ormel, 2000) and mental disorders (Rhode, Lewinsohn and Seeley, 1990; Broadhead, Blazer, George and Tse, 1990; Tweed, 1993; Coryell, Scheftner, Keller, Endicott, Maser and Klerman, 1993; Ormel et al., 1994; Ormel et al., 1998; Bijl and Ravelli, 2000). Such results have led to an interest among health policy researchers in the possibility that expanded outreach and guideline-concordant treatment of impairing chronic disorders might represent an investment opportunity for employers (Kessler, Greenberg, Mickelson, Meneades and Wang, 2001) as well as for governments (Murray and Lopez, 1996). However, not all mental disorders have been studied in this way. The WMH-CIDI assessment of internal impairment for each mental disorder and for one random chronic physical condition allows this to be done.

The WMH-CIDI assessment of external impairment has the potential to be even more important in this regard. A central limitation of the existing literature on the role impairments of chronic conditions is the lack of attention to the co-occurrence of multiple disorders in the same patient. Many people with chronic disorders suffer from more than one disorder (Dewa and Lin, 2000). Pure disorders are, in general, less impairing than co-occurring disorders in clinical samples (Ormel et al., 1994). The co-occurrence of mental disorders with chronic physical disorders is of special importance in this regard, as strong patterns of co-occurrence with mental disorders have been found for a number of commonly occurring physical disorders both in general population samples (Neeleman, Ormel and Bijl, 2001) and in primary care samples (Berardi, Berti Ceroni, Leggieri, Rucci, Üstün and Ferrari, 1999).

Clinical studies have found excess impairment associated with co-occurring mental disorders among people with chronic physical disorders (Sullivan, LaCroix, Russo and Walker, 2001). As efforts increase to rationalize the allocation of healthcare resources guided by the criteria of evidence-based medicine, consideration of the role played by co-occurring mental disorders in causing impairment among patients with chronic physical disorders becomes all the more important. The inclusion of the external impairment in the WMH-CIDI makes it possible to carry out such analyses in general population samples by using information about mental disorders, physical disorders, and their comorbidities to predict external impairment.

Part I and Part II diagnoses

The WMH-CIDI is quite a long instrument, with an average administration time of approximately 2 hours for the full interview. This long administration time can create practical complications, the most important of them being that it is often necessary to administer the interview in two sessions. In order to address this length problem, a case-control approach is used in developing the WMH-CIDI whereby a subsample of respondents who complete the first half of the interview (Part I), which includes all core diagnostic assessments, and who report having no lifetime history of disorder, are terminated at this mid-point of the interview. All respondents who meet criteria for any lifetime mental disorders in the Part I interview,

in comparison, are retained in the second half of the interview (Part II) along with a probability subsample of non-cases. The default value for the non-case probability of selection is 25%, although this can be changed depending on the interests of the investigator. This sampling fraction leads to Part II samples typically retaining between 33% and 67% of all Part I respondents. This case-control subsampling fraction yields a high ratio of controls to cases for all but the most prevalent disorders. Statistical power analysis show that increasing the number of controls to cases above these levels yields very little improvement in power (Schlesselman, 1982). As the main purposes of carrying out a WMH-CIDI survey are to estimate prevalence and correlates of mental disorders, this result implies that the subsampling of non-cases into Part II retains most of the efficiency of the full sample for central analyses while substantially reducing field costs.

Once the Part II subsampling approach is in place, it becomes clear that average interview length is reduced whenever a section is moved from Part I to Part II. Several important, but lengthy, diagnostic sections in the WMH-CIDI were included in Part II for this reason. These include the assessments of post-traumatic stress disorder, obsessive-compulsive disorder, and non-affective psychosis. It should be noted that the high comorbidity of these Part II disorders with the disorders assessed in Part I means that the great majority of respondents with these disorders are selected into the Part II sample, leading to only a small loss of information about the Part II disorders by placing them in Part II rather than in Part I. In addition, disorders that are included in the WMH-CIDI for exploratory purposes are all placed in Part II. Included here are eating disorders, neurasthenia, nicotine dependence, pathological gambling, premenstrual disorder, and a screen for personality disorders.

Options for additional subsampling of the assessment of these exploratory disorders within the Part II sample (for example, only a random 50% of Part II respondents receive the assessment of pathological gambling) are built into the WMH-CIDI skip logic for investigators who vary in their level of interest in these exploratory disorders. In addition, a series of four diagnostic sections are included in Part II for retrospectively reported childhood and adolescent disorders. These sections, which are modelled on

those developed by Lee Robins for the DSM-IV version of the DIS, include assessments of Attention-Deficit/Hyperactivity Disorder, Conduct Disorder, Oppositional-Defiant Disorder, and Separation Anxiety Disorder.

Expansions of other WMH-CIDI sections

As noted previously, the main reason for developing the WMH-CIDI is to expand the instrument beyond its initial focus on diagnoses to include assessments of risk factors, consequences, and treatment. The WMH-CIDI includes 14 sections of this sort. Six sections assess socio-demographics (employment, finances, marriage, children, adult and childhood demographics for the Part I sample, adult demographics for the Part II sample). Two sections assess treatment (services, pharmacoepidemiology). The other six sections assess external impairment, chronic conditions, non-specific psychological distress, social networks, family burden, and childhood experiences. We already commented on the external impairment and chronic conditions sections. A few comments are in order about several of the other sections.

Socio-demographics

Basic information about socio-demographic variables – such as age, sex, race, education, marital status, and employment status – is included in all community surveys. For the most part, though, this information is cross-sectional – it assesses the respondent's current status on these variables rather than his or her history. This is fine for ascribed socio-economic characteristics (such as sex and race), which do not change over time, but it misses important information about the dynamics of achieved statuses, such as marital status and employment status, which change over time. This loss of information can be important if dynamic information is relevant to mental health. Given the focus of the WMH-CIDI diagnostic sections on lifetime course, we feel that it is important to include dynamic information about achieved statuses in the interview schedule. As a result, separate sections of the interview are devoted to the respondent's history in each of the three main areas of achieved social status – employment, marriage, and childbearing.

The details of the assessment differ across the three sections, but the basic approach is the same.

We begin by asking about timing of initial entry into roles (for example ages at first dating, first marriage, first employment, first sexual intercourse, first becoming pregnant (females) or causing a woman to become pregnant (males), first having an abortion, first giving birth). We then ask about role history (for example, age at onset and duration of each marriage, age of each child, stability of employment history). Information about current role incumbency (for example, current employment status, current marital status, number and ages of children, which children live with the respondent), which is the focus of the assessment in most surveys, is only the final part of the assessments. The exception is Part I respondents terminated before the Part II interview, who are administered a brief socio-demographic battery that focuses only on current status.

The additional socio-demographic sections – childhood demographics, adult demographics, and finances – also obtain much more detail about these areas than in most other surveys. The section on childhood demographics asks about age of parents when the respondent was born, size of sib-ship and birth order, marital status of parents, nativity, number of generations in the country among natives, age at immigration among the foreign born, country of origins for people who were originally from another country, native tongue, education, childhood religion and religiosity, urbanicity of childhood residence, and stability of childhood residence. The section on adult demographics asks about whether parents are living or dead, age and cause of death of each deceased parent, race-ethnicity, subjective closeness of racial-ethnic identification, citizenship, religious preference, religiosity, amount of time during adulthood when the respondent was in a jail or prison or correctional facility, amount of time homeless, amount of time institutionalized in a hospital or nursing home, and current subjective social class position. The section on finances asks both about objective finances and subjective financial stress. With regard to objective finances, information is obtained both on income and assets. Income information is obtained for the entire household broken down by income of the respondent, the respondent's spouse, other family members, income from government assistance programmes, and other income. This disaggregation of income is very useful in analyses of social class and mental illness, where it

is possible to distinguish associations that might be due to direct selection (respondent income), assortative mating (spouse income), and other sources.

Treatment and pharmacoepidemiology

Treatment, like impairment, is assessed in the WMH-CIDI both internally and externally. In the internal assessment, respondents who meet criteria for a particular disorder are asked at the end of the diagnostic section whether they ever sought professional treatment for that disorder and, if so, at what age they first sought this treatment. They are also asked if they ever obtained treatment that they considered helpful for the disorder and, if so, how many different professionals they had to see before they received helpful treatment. Respondents who say they never received helpful treatment are asked how many professionals they ever saw for the disorder. This kind of information, when coupled with information on age at onset of the disorder, can be used to study patterns and predictors of delays in initial treatment contact after first onset of a mental disorder.

Analyses of this sort in epidemiological samples consistently find pervasive delays in initial treatment contact after first onset of a mental disorder that are inversely related to age at onset, cohort, and illness severity (Kessler, Olfson and Berglund, 1998; Olfson, Kessler, Berglund and Lin, 1998). It is also unclear whether patient reports of being helped would be confirmed in objective evaluations. Nonetheless, patient perception, even if not entirely accurate, has to be considered an important dimension of treatment effectiveness that has not previously been seriously considered in psychiatric epidemiological surveys.

In the external assessment of treatment, the WMH-CIDI asks respondents about ever having treatment for problems with their emotions or mental health. Questions about inpatient treatment include asking about lifetime hospitalization, age of first hospitalization, number of lifetime hospitalizations, amount of time spent in hospitals for these problems over the life course, and hospitalization in the past 12 months. Questions about outpatient treatment include asking about treatment from each of a wide range of professionals. For each type of professional seen, information is recorded on age of first receiving treatment and age of most recent treatment. For those who received treatment in the past 12 months, information is obtained on number

of visits with each type of professional, average duration of time with the professional, whether the respondent is still in treatment and, if not, whether the termination of treatment occurred because the respondent completed the course of treatment or quit. Summary questions are then asked of all respondents who were in any type of 12-month treatment about all the money spent out of pocket for treatment with all professionals over that interval of time, reasons for seeking treatment, and reasons for terminating treatment among those who did terminate. Respondents who did not receive any treatment in the past 12 months are asked whether they ever felt that they might need professional help for their emotions or mental health in the past 12 months. If so, they are asked about reasons for not seeking professional help.

A separate pharmacoepidemiology section asks about the use of prescription and non-prescription medications in the past 12 months for 'problems with your emotions, nerves, mental health, substance use, energy, concentration, sleep, or ability to cope with stress'. An exhaustive list of prescription medications is provided as a visual aid in answering these questions. Interviewers are instructed to use motivational probes to encourage respondents to think carefully and exhaustively to list all medications taken during the recall period. Interviewers are also instructed to have respondents get their medicine bottles if they still have them in order to copy down information about the name of the medicine and the recommended dose. For each medicine taken, questions are then asked about number of days taken out of the past 365, dose, whether the medicine is taken under the supervision of a doctor. If not, information is collected about where the respondent obtains the medicine. If taken under the supervision of a doctor, a question is asked whether the doctor is a psychiatrist, some other mental health specialist, a primary care doctor, or some other kind of doctor. If taken under professional supervision, a question is asked how about the respondent failed to take the medicine at the recommended dose and times. Finally, questions are asked about whether the medicine is still being taken and, if not, reasons for terminating use, including side effects.

Non-specific psychological distress

The Part II interview includes screening scales about

the frequency of non-specific psychological distress during the 30 days before the interview and in the worst month of the past year. These scales were developed originally for use in the NHIS in order to screen for Serious Mental Illness (SMI) (defined as any DSM anxiety or mood or psychotic disorder associated with a Global Assessment of Functioning score less than 60). They have subsequently been adopted not only by the NHIS, but also by a number of other large ongoing national government health surveys in the US as well as by the Australian and Canadian governments for use in their ongoing health surveys. The 10 questions in the scales were selected from a large item pool aimed at sensitively measuring the first principal factor of non-specific psychological distress, which is consistently found in community surveys of distress, in the clinically significant range of its distribution (Furukawa, Andrews, Slade and Kessler, 2003).

A validation study carried out in a community sample showed that the 10-question scale does an excellent job of screening for SMI (Kessler, Barker, Colpe, Epstein, Gfroerer, Hiripi, Howes, Normand, Manderscheid, Walters and Zaslavsky, 2003). However, the sensitivity and specificity of the scale might vary across populations making it useful to include the scale in WMH-CIDI surveys so that new calibrations can be made across many different populations. When accurate rules of this sort exist, this brief screening scale can be very useful as an inexpensive mental health needs surveillance tool in ongoing general-purpose tracking surveys. Calibration rules and software for transforming scores on this scale into individual-level predicted probabilities of SMI and other global measures of disorder are being posted as they are developed, not only for the US but also for all countries in the WMH Survey Initiative.

Family burden

Although the questions on internal and external impairment do an excellent job of assessing the ways in which mental disorders affect the people who have these disorders, the effects on the families of the mentally ill are ignored. This is a major gap as there is clear evidence that mental disorders can create enormous family burdens (Saunders, 2003; Tsang, Tam, Chan and Chang, 2003). In an effort to correct this situation, a separate section was developed for the WMH-CIDI on family burden. Unlike all other sections of the interview, this section treats the respondent as the 'family member' whose burden is being assessed rather than as the person whose mental health is being assessed. This was based on concerns that respondents would be unable to provide accurate informant information on the burdens their disorders imposed on their loved ones. An additional virtue of this method is that it provides an easy way to integrate information on burden into population estimates, something that would be extremely difficult to do without enumerating networks if the respondent was treated as the focal respondent whose illness affected many different family members rather than as a representative family member who might be burdened by the illness of any number of family members.

The logic of the section requires us to begin by defining a network of first-degree relatives. This is done by asking the respondent how many living first-degree relatives he has, separately reporting the number of parents, siblings, children, and whether or not he has a spouse. The respondent is then told that we have a few questions about the health problems of these individuals. After enumerating the network in this way, the respondent is asked if any of these individuals has any of 12 serious health problems. These include the following, in the order they are asked: cancer; serious heart problems; a serious memory problem, like senility or dementia; mental retardation; a permanent physical disability, like blindness or paralysis; any other serious chronic physical illness; alcohol or drug problems; serious depression; serious anxiety; schizophrenia or psychosis; manic-depression; or any other serious chronic mental illness. Note that the threshold of problems is set high (for example, 'serious depression' rather than 'depression') and that open-ended questions are included about other serious chronic physical and mental illnesses at the end of the lists of explicit physical and mental disorders. The open ends are put at the end so that the earlier explicit disorders can provide a context for defining the severity level we want respondents to be thinking about when responding to the open-ended questions.

Respondents who report having any first-degree relatives with any of the 12 health problems are then asked a series of questions about how these problems, taken as a whole, affect the life of the respondent. Included here are questions about the respondent

helping to wash or dress the ill person, helping with practical things (for example, paperwork, housework, local transportation, and taking medications), and spending more time with the ill person(s) or giving them more emotional support than they would if the illness(es) were not present. The respondent then estimates the number of hours per week he or she spends doing things related to the health problems of these family members and the amount of money he spends per month because of these problems. Emotional effects of the problems on the respondent are then briefly assessed (for example, extent of worry, anxiety, depression, and embarrassment caused by the problems). Finally, the Sheehan Disability Scales are used to have the respondent rate the extent to which family member health problems interfere with his or her functioning in the same four areas of functioning used in the assessment of internal impairment (household duties, employment, social life, and close personal relationships).

The use of the Sheehan scales makes it possible to compare the effects on functioning of the respondent's own illnesses compared to the illnesses of family members. More generally, by creating 48 separate variables (for each of the 12 illnesses for each of the four types of first-degree relatives, using counts to deal with the situation where more than one family member of a given type has a particular type of illness), regression analysis can be used to examine the relative effects of different types of illnesses on the various dimensions of family burden included in the assessment. It is also possible, using this same analysis approach, to study whether particular aspects of burden are greater for female than male focal respondents exposed to the same profile of family illness, to examine how burden changes when the structural relationship between the focal respondent and the ill person changes (for example, the illness of a child versus of a parent), and to evaluate the effects of network illness comorbidity on respondent burden. As far as we are aware, none of these has up to now been the subject of systematic analysis across a wide range of illness categories considered together in a large-scale community survey, although each of them has long been studied in small-scale focused studies of patients and their families (for example, Chakrabarti, Kulhara and Verma, 1993; Rupp and Keith, 1993). The inclusion of the family burden section in the WMH-CIDI

makes it possible to carry out this systematic kind of analysis.

Childhood experiences

The final section in the WMH-CIDI that we want to mention is the section on childhood experiences. The WMH collaborators agreed early on that the WMH surveys were uniquely positioned to study the lifetime effects of traumatic life experiences, with a special emphasis on the long-term effects of childhood adversities. As a result, a fairly extensive series of questions is included in the WMH-CIDI about childhood experiences. The questions about childhood traumatic events are placed in the trauma checklist within the PTSD section. The remaining questions are included in a separate section on childhood experiences. This section begins by asking whether the respondent lived with both of his or her biological parents until age 16 and, if not, to explain his or her living situation up to that age. The nature and age at each important transition obtained in response to this question are recorded for such events as death of a parent, parental divorce, adoption, and the like. Respondents who report living with both parents up to age 16 are then asked whether a parent was ever away from home for six months or longer due to such things as hospitalization, imprisonment, or military service. The respondent's age at, and the duration of, each such event are recorded. A question is then asked about whether the respondent was ever away from home for 6 months or longer due to such things as hospitalization, boarding school, foster care, or residential treatment. The respondent's age at, and the duration of, each such event are recorded.

The remainder of the section focuses on respondent reports about their mothers and fathers or, in the absence of a mother or father during their childhood, the man and woman who served as the equivalents of mother and father. Focusing on the main parent figure, respondents are asked their biological relationship to this individual (for example, grandfather or step-father) as well as the closeness and warmth of their interpersonal relationships during the years the respondent was growing up. A modified version of the Parental Bonding Instrument (Parker, 1989) is used to classify parent-child relationships as either authoritarian, authoritative, overprotective, or neglectful. A modified version of the Conflict Tactics Scale (Straus,

Hamby, Finkelhor, Moore and Runyan, 1998) is used to assess the frequency and intensity of parental violence towards the respondent during the respondent's childhood. Questions are also asked about neglect and sexual abuse. Information is then obtained on parent education, employment status, occupation if they were employed, and the stability of their employment during the respondent's childhood. Finally, a modified version of the Family History Research Diagnostic Criteria Interview (Andreasen, Endicott, Spitzer and Winokur, 1977) is used to assess parental psychopathology during the respondent's childhood. Separate assessments are made here of parental depression, panic disorder, GAD, substance-use disorder, and antisocial personality disorder.

Overview and future directions

As noted in the introduction, the CIDI was originally developed by WHO to be a tool that could help coordinate the efforts of psychiatric epidemiologists around the world to carry out community surveys in which results could be directly compared and cumulated by virtue of using the same instrument. The WMH-CIDI continues in this tradition by refining the diagnostic assessments in the original CIDI and by adding sections that examine risk factors, consequences, and treatment. Like the original CIDI, the WMH-CIDI was designed to generate diagnoses using the definitions and criteria of both the ICD and DSM systems (ICD-10 and DSM-IV). In addition, like the original CIDI, the WMH-CIDI has been translated into a number of languages using the standard WHO translation and back-translation protocol.

An important WMH-CIDI development is the establishment of an explicit protocol for modifying the instrument. In the past, reluctance on the part of the WHO CIDI Advisory Committee (WHO CIDI-AC) to work with investigators who proposed modifications led to idiosyncratic changes made by individual users in different surveys that reduced comparability across studies. The new protocol for modifying the WMH-CIDI calls for users who want to modify diagnostic questions to include both the original WMH-CIDI questions and the proposed new questions in their modified version of the instrument and to carry out blind clinical follow-up interviews in a stratified probability sample of

concordant and discordant cases in order to evaluate whether the new questions increase consistency of WMH-CIDI diagnoses with clinical diagnoses. In cases where the old and new questions cannot logically be included in the same instrument, a split ballot approach is stipulated in which random subsamples receive one or the other in the same study.

Stipulations for the design, instrumentation, and quality control of WMH-CIDI clinical reappraisal studies have been established to guarantee endorsement of results by the WHO CIDI-AC. Suggestions for such methodological studies and results of these studies will be posted on the WMH-CIDI Web page along with author attributions. Replication of positive results in a second endorsed methodological study will lead to proposed changes being adopted in the next revision of the WMH-CIDI. A similar system of making proposed modifications and expansions of the CIDI in ways that do not change the diagnostic questions will also be posted on the WMH-CIDI Web page in order to create a library of potentially useful alternative questions for future users. Included here, for example, might be expanded questions about childhood adversity, a new section on coping, or more elaborate questions about the nature of specific fears aimed at subtyping specific phobias. As with other proposed modifications and expansions of the instrument, author attributions will be included with each of these postings.

Pincus (University of Pittsburgh), Greg Simon (Group Health Cooperative), Michael Von Korff (Group Health Cooperative), Philip Wang (Harvard Medical School), Kenneth Wells (UCLA), Elaine Wethington (Cornell University) and Hans-Ulrich Wittchen (Institute of Clinical Psychology, Technical University Dresden and Max Planck Institute of Psychiatry). The authors appreciate the helpful comments on earlier drafts of Robert Belli, Beth-Ellen Pennell, Norbert Schwarz, and Hans-Ulrich Wittchen. Portions of this paper were previously published in Kessler et al. (2000) and are used here with the permission of the publisher, Lawrence Erlbaum Associates. A complete list of NCS publications and the full text of all NCS-R instruments can be found at http://www.hcp.med.harvard.edu/ncs. Send correspondence to NCS@hcp.med.harvard.edu.

The development of the WMH-CIDI depended on the efforts of a great many individuals. Clinical research experts who worked with us on specific diagnostic sections include the following: Jules Angst (major depression, mania), Gavin Andrews (anxiety disorders), James Anthony (substance disorders), Naomi Breslau (post-traumatic stress disorder), Emil Coccaro (intermittent explosive disorder), Ian Falloon (non-affective psychosis), Abby Fyer (agoraphobia, specific phobia), Josep Maria Haro (family burden), Richard Heimberg (social phobia), David Katzelnick (obsessive-compulsive disorder), Martin Keller (generalized anxiety disorder), Mark Lenzenweger (personality disorders), Armand Lorenger (personality disorders), Bruce Lydiard (irritable bowel syndrome), Susan McElroy (intermittent explosive disorder), Kathleen Merikangas (major depression, mania, nicotine dependence), Dan Mroczek (non-specific psychological distress), Tom Roth (sleep disorders), John Rush (major depresson), Marc Schuckit (substance use disorders), Kathy Shear (panic disorder, separation anxiety disorder), Derrick Silove (separation anxiety disorder), Dan Stein (post-traumatic stress disorder, obsessive-compulsive disorder), Mauricio Tohen (bipolar disorder), Michael von Korff (chronic pain disorder), Lynn Wallisch (pathological gambling), Philip Wang (services, pharmacoepidemiology), Ken Winters (pathological gambling), and Hans-Ulrich Wittchen (anxiety disorders). Survey methodology experts who worked with us on the various question wording studies and refinements discussed in the first half of the paper include Robert Belli, Charles Cannell, Barbel Knauper, Beth-Ellen Pennell, and Norbert Schwarz. The WMH-CIDI built of the pioneering work of Lee Robins in developing the Diagnostic Interview Schedule and the original CIDI and of Hans-Ulrich Wittchen in developing a number of important CIDI innovations in his M-CIDI. Neither the original DIS nor the CIDI and its extensions would have been possible without the vision of Darrel Regier, who launched the programme of research that led to the DIS and the CIDI and subsequently the WMH-CIDI.

References

Aguilar-Gaxiola S, Alegria M, Andrade L, Bijl RV, Caraveo-Anduaga JJ, DeWit DJ, Kolody B, Kessler RC, Üstün TB, Vega WA, Wittchen HU. The International Consortium in Psychiatric Epidemiology. In E Dragomirecká, A Palcová and H Papezová eds. Social Psychiatry in Changing Times. Prague, Czech Republic: Prague Psychiatric Center, 2000.

Alegria M, Kessler RC, Bijl R, Lin E, Heeringa S, Takeuchi DT, Kolody B. Comparing mental health service use data across countries. In G Andrews ed. Unmet Need in Mental Health Service Delivery. Cambridge, United Kingdom: Cambridge University Press, 2000, 97–118.

Andreasen NC, Endicott J, Spitzer RL, Winokur G. The family history method using diagnostic criteria. Reliability and validity. Arch Gen Psychiatry 1977; 34: 1229–35.

Angst J, Merikangas K, Scheidegger P, Wicki W. Recurrent brief depression: a new subtype of affective disorder. J Affect Dis (Amsterdam) 1990; 19: 87–98.

Belli RF. Color blend retrievals: compromise memories or deliberate compromise responses? Mem Cognit 1988; 16: 314–26.

Belson WA. The Design and Understanding of Survey Questions. Aldershot: Gower, 1981.

Benjamin J, Ebstein RP, Lesch KP. Genes for personality traits: implications for psychopathology. Int J Neuropsychopharmacol 1998; 1: 153–68.

Berardi D, Berti Ceroni G, Leggieri G, Rucci P, Üstün TB, Ferrari G. Mental, physical and functional status in primary care attenders. Int J Psychiatry Med 1999; 29: 133–48.

Bijl RV, de Graaf R, Hiripi E, Kessler RC, Kohn R, Offord DR, Üstün TB, Vicente B, Vollebergh WA, Walters EE, Wittchen HU. The prevalence of treated and untreated mental disorders in five countries. Health Aff (Millwood) 2003; 22: 122–33.

Bijl RV, Ravelli A. Current and residual functional disability associated with psychopathology: findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Psychol Med 2000; 30: 657–68.

Breslau N, Davis GC. DSM-III generalized anxiety disorder: an empirical investigation of more stringent criteria. Psychiatry Res 1985; 15: 231–8.

Brewer WF. What is autobiographical memory? In DC Rubin ed. Autobiographical Memory. New York: Cambridge University Press, 1986, 25–49.

Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA 1990; 264: 2524–8.

Brown GW. Life events and measurement. In GW Brown, TO Harris eds. Life Events and Illness. New York: Guilford Press, 1989, 3–45.

Brugha TS, Bebbington PE, Jenkins R. A difference that matters: comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. Psychol Med 1999; 29: 1013–20.

Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE. A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). Psychol Med 2001; 31: 1001–13.

Cannell CF, Fowler FJ, Jr., Marquis KH. The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. Vital Health Stat 1 1968; 2: 1–65.

Cannell CF, Miller PV, Oksenberg L. Research on interviewing techniques. In S Leinhardt ed. Sociological Methodology. San Francisco: Jossey-Bass, 1981, 389–437.

Centers R. A laboratory adaptation of the conversational procedure for the conditioning of verbal operants. Journal of Abnormal and Social Psychology 1964; 67: 334–9.

Chakrabarti S, Kulhara P, Verma SK. The pattern of burden in families of neurotic patients. Soc Psychiatry Psychiatr Epidemiol 1993; 28: 172–7.

Clark HH, Schober MF. Asking questions and influencing answers. In JM Tanur ed. Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation, 1992, 15–48.

Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960; 20: 37–46.

Coryell W, Scheftner W, Keller M, Endicott J, Maser J and Klerman GL. The enduring psychosocial consequences of mania and depression. Am J Psychiatry 1993; 150: 720–7.

Dewa CS, Lin E. Chronic physical illness, psychiatric disorder and disability in the workplace. Soc Sci Med 2000; 51: 41–50.

Eaton WW, Badawi M, Melton B. Prodromes and precursors: epidemiologic data for primary prevention of disorders with slow onset. Am J Psychiatry 1995; 152: 967–72.

Eaton WW, Kessler RC, Wittchen HU, Magee WJ. Panic and panic disorder in the United States. Am J Psychiatry 1994; 151: 413–20.

Eaton WW, Neufeld K, Chen LS, Cai G. A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule for clinical assessment in neuropsychiatry in the Baltimore Epidemiologic Catchment Area follow-up. Arch Gen Psychiatry 2000: 217–22.

Edwards WS, Winn DM, Kurlantzick V, Sheridan S, Berk ML, Retchin S, Collins JG. Evaluation of National Health Interview Survey Diagnostic Reporting. National Center for Health Statistics. Vital Health Stat 2 1994; 120: 1–116.

Furukawa TA, Andrews G, Slade T, Kessler RC. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. Psychol Med 2003; 33: 357–62.

Gilman S. Disease and Representation: Images of Illness from Madness to AIDS. Ithaca NY: Cornell University Press, 1988.

Gross R, Bentur N, Elhayany A, Sherf M, Epstein L. The validity of self-reports on chronic disease: characteristics of underreporters and implications for the planning of services. Public Health Rev 1996; 24: 167–82.

Halabi S, Zurayk H, Awaida R, Darwish M, Saab B. Reliability and validity of self and proxy reporting of morbidity data: a case study from Beirut, Lebanon. Int J Epidemiol 1992; 21: 607–12.

Hays RD, Wells KB, Sherbourne CD, Rogers W, Spritzer K. Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. Arch Gen Psychiatry 1995; 52: 11–19.

Heliovaara M, Aromaa A, Klaukka T, Knekt P, Joukamaa M, Impivaara O. Reliability and validity of interview data on chronic diseases. The Mini-Finland Health Survey. J Clin Epidemiol 1993; 46: 181–91.

Jobe JB, White AA, Kelley CL, Mingay DL, Sanchez MJ, Loftus EF. Recall strategies and memory for health care visits. Milbank Q 1990; 68: 171–89.

Judd LL, Paulus MP, Wells KB, Rapaport MN. Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population. Am J Psychiatry 1996; 153: 1411–17.

Kempen GI, Ormel J, Brilman EI, Relyveld J. Adaptive responses among Dutch elderly: the impact of eight chronic conditions on health-related quality of life. Am J Public Health 1997; 87: 38–44.

Kempen GI, Sanderman R, Miedema I, Mayboom-de JB, Ormel J. Functional decline after congestive heart failure and acute myocardial infarction and the impact of psychological attributes. A prospective study. Quality Life Res 2000; 9: 439–50.

Kendell RE. Five criteria for an improved taxonomy of mental disorders. In JE Helzer, JJ Hudziak eds. Defining Psychopathology in the 21st Century DSM-V and Beyond. Washington DC: American Psychiatric Publishing, 2002, 3–17.

Kessler R. The World Health Organization International Consortium in Psychiatric Epidemiology (ICPE): Initial work and future directions – the NAPE lecture 1998. Acta Psychiatr Scand 1999; 99: 2–9.

Kessler RC, Barber C, Beck A, Berglund PA, Cleary PD, McKenas D, Pronk N, Simon G, Stang P, Üstün TB, Wang PS. The World Health Organization Health and Work Performance Questionnaire (HPQ). J Occup Environ Med 2003; 45: 156–74.

Kessler RC, Barker PR, Colpe LJ, Epstein A, Gfroerer JC,

Hiripi E, Howes MJ, Normand S-LT, Manderscheid RW, Walters EE, Zaslavsky AM. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003; 60: 184–9.

Kessler RC, Greenberg PE, Mickelson KD, Meneades LM, Wang PS. The effects of chronic medical conditions on work loss and work cutback. J Occup Environ Med 2001; 43: 218–25.

Kessler RC, Merikangas K. The National Comorbidity Survey Replication (NCS-R). International Journal of Methods in Psychiatric Research 2004; (this issue).

Kessler RC, Mroczek DK and Belli RF. Retrospective adult assessment of childhood psychopathology. In D Shaffer, CP Lucas, JE Richters eds. Diagnostic Assessment in Child and Adolescent Psychopathology. New York: Guilford Press, 1999, 256–84.

Kessler RC, Olfson M, Berglund PA. Patterns and predictors of treatment contact after first onset of psychiatric disorders. Am J Psychiatry 1998; 155: 62–9.

Kessler RC, Üstün TB. The World Health Organization World Mental Health 2000 Initiative. Hospital Management International 2000: 195–6.

Kessler RC, Wittchen H-U, Abelson JM, McGonagle K, Schwarz N, Kendler KS, Knauper B, Zhao S. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. International Journal of Methods in Psychiatric Research 1998; 7: 33–55.

Kessler RC, Wittchen HU, Abelson JM, Zhao S. Methodological issues in assessing psychiatric disorder with self-reports. In AA Stone, JS Turrkan, CA Bachrach, JB Jobe, HS Kurtzman, VS Cain eds. The Science of Self-Report: Implications for Research and Practice. Mahwah NJ: Lawrence Erlbaum Associates, 2000, 229–25.

Kriegsman DM, Penninx BW, van Eijk JT, Boeke AJ, Deeg DJ. Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. A study on the accuracy of patients' self-reports and on determinants of inaccuracy. J Clin Epidemiol 1996; 49: 1407–17.

Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med 1997; 27: 93–105.

Leventhal H, Nerenz D, Steele DJ. Illness representations and coping with health threats. In A Baum, SE Taylor, JE Singer eds. Handbook of Psychology and Health 4. Hillsdale NJ: Erlbaum, 1984, 219–52.

Lipton RB, Dodick D, Sadovsky R, Kolodner K, Endicott J, Hettiarachchi J, Harrison W. A self-administered screener for migraine in primary care: The ID Migraine(TM) validation study. Neurology 2003; 61: 375–82.

Loftus EF, Palmer JC. Reconstruction of automobile destructions: an example of the integration between language and memory. Journal of Verbal Language and Verbal Behavior 1974; 13: 585–9.

Mackenbach JP, Looman CW, Van der Meer JB. Differences in the misreporting of chronic conditions, by level of education: the effect on inequalities in prevalence rates. Am J Public Health 1996; 86: 706–11.

Maier W, Gansicke M, Freyberger HJ, Linz M, Heun R, Lecrubier Y. Generalized anxiety disorder (ICD-10) in primary care from a cross-cultural perspective: a valid diagnostic entity? Acta Psychiatr Scand 2000; 101: 29–36.

Markus H, Zajonc RB. The cognitive perspective in social psychology. In G Lindzey, E Aronson eds. The Handbook of Social Psychology. New York: Random House, 1985, 137–230.

Marlatt GA. Task structure and the experimental modification of verbal behavior. Psychol Bull 1972; 78: 335–50.

Marquis KH, Cannell CF. A Study of Interviewer-Respondent Interaction in the Urban Employment. Ann Arbor MI: Survey Research Center, University of Michigan, 1969.

Means B, Loftus EF. When personal history repeats itself: decomposing memories for recurring events. Applied Cognitive Psychology 1991; 5: 297–318.

Mechanic D. Is the prevalence of mental disorders a good measure of the need for services? Health Aff (Millwood) 2003; 22: 8–20.

Menon A. Judgments of behavioral frequencies: memory search and retrieval strategies. In N Schwartz, S Sudman eds. Autobiographical Memory and the Validity of Retrospective Reports. New York: Springer-Verlag, 1994, 161–72.

Merikangas KR, Zhang H, Avenevoli S, Acharyya S, Neuenschwander M, Angst J. Longitudinal trajectories of depression and anxiety in a prospective community study: the Zurich Cohort Study. Arch Gen Psychiatry 2003; 60: 993–1000.

Miller PV, Cannell CF. Communicating measurement objectives in the survey interview. In DM Hirsch, PV Miller, FG Kline eds. Strategies for Communication Research 6. Beverly Hills CA: Sage, 1977.

Murray CJL, Lopez AD. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020. Cambridge MA: Harvard University Press, 1996.

Narrow WE, Rae DS, Robins LN, Regier DA. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile two surveys' estimates. Arch Gen Psychiatry 2002; 59: 115–23.

National Advisory Mental Health Council. Health care reform for Americans with severe mental illnesses: report of the National Advisory Mental Health Council. Am J Psychiatry 1993; 150: 1447–65.

National Center for Health Statistics DoDS. 2003 National Health Interview Survey (NHIS). Hyattsville MD: US Department of Health and Human Services, CDC, 2003.

Neeleman J, Ormel J, Bijl RV. The distribution of psychiatric and somatic ill-health: associations with personality and socioeconomic status. Psychosom Med 2001; 63: 239–47.

Oksenberg L, Cannell CF, Kanton G. New strategies for pretesting survey questions. Journal of Official Statistics 1991; 7: 349–65.

Oksenberg L, Vinokur A, Cannell CF. The effects of instructions, commitment and feedback on reporting in personal interviews. In CF Cannell, L Oksenberg, JM Converse eds. Experiments in Interviewing Techniques. DHEW Publication No. (HRA) 78-3204. Washington DC: Department of Health, Education, and Welfare, 1979a, 133–99.

Oksenberg L, Vinokur A, Cannell CF. Effects of commitment to being a good respondent on interview performance. In CF Cannell, L Oksenberg, JM Converse eds. Experiments in Interviewing Techniques. DHEW Publication No. (HRA) 78-3204. Washington DC: Department of Health, Education, and Welfare, 1979b, 74–108.

Olfson M, Kessler RC, Berglund PA and Lin E. Psychiatric disorder onset and first treatment contact in the United States and Ontario. Am J Psychiatry 1998; 155: 1415–22.

Ormel J, Kempen GI, Deeg DJ, Brilman EI, Van Sonderen E, Relyveld J. Functioning, well-being, and health perception in late middle-aged and older people: comparing the effects of depressive symptoms and chronic medical conditions. J Am Geriatr Soc 1998; 46: 39–48.

Ormel J, Von Korff N, Üstün TB, Pini S, Korten A, Oldehinkel T. Common mental disorders and disability across cultures: results from the WHO collaborative study on psychological problems in general health care. JAMA 1994; 272: 1741–8.

Parker G. The Parental Bonding Instrument: psychometric properties reviewed. Psychiatr Dev 1989; 7: 317–35.

Pearson RW, Ross M, Dawes RM. Personal recall and the limits of retrospective questions in surveys. In JM Tanur ed. Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation, 1992, 65–94.

Penninx BW, Beekman AT, Ormel J, Kriegsman DM, Boeke AJ, Van Eijk JT, Deeg DJ. Psychological status among elderly people with chronic diseases: does type of disease play a part? J Psychosom Res 1996; 40: 521–34.

Pincus HA, Zarin DA, First M. 'Clinical Significance' and DSM-IV. Arch Gen Psychiatry 1998; 55: 1145.

Preisig M, Merikangas KR and Angst J. Clinical significance and comorbidity of subthreshold depression and anxiety in the community. Acta Psychiatr Scand 2001; 104: 96–103.

Regier DA. Community diagnosis counts [Commentary]. Arch Gen Psychiatry 2000; 57: 223–4.

Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS. Limitations of diagnostic criteria and assessment instruments for mental disorders. Implications for research and policy. Arch Gen Psychiatry 1998; 55: 109–15.

Regier DA, Narrow WE. Defining clinically significant psychopathology with epidemiologic data. In JE Helzer, JJ Hudziak eds. Defining Psychopathology in the 21st Century: DSM-V and Beyond. Washington DC: American Psychiatric Publishing, 2002, 19–30.

Rehm J, Üstün TB, Saxena S, Nelson CB, Chatterji S, Ivis F, Adlaf E. On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. International Journal of Methods in Psychiatric Research 1999; 8: 110–23.

Rhode P, Lewinsohn P, Seeley J. Are people changed by the experience of having an episode of depression? A further test of the scar hypotheses. J Abnorm Psychol 1990; 99: 264–71.

Robins LN, Helzer JE, Croughan JL, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. Arch Gen Psychiatry 1981; 38: 381–9.

Robins LN, Regier DA. Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. New York: The Free Press, 1991.

Robins JN, Wing J, Wittchen HU, Helzer JE, Babor TE, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH. The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 1988; 45: 1069–77.

Rupp A, Keith SJ. The costs of schizophrenia. Assessing the burden. Psychiatr Clin North Am 1993; 16: 413–23.

Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996; 26: 477–86.

Saunders JC. Families living with severe mental illness: a literature review. Issues Ment Health Nurs 2003; 24: 175–98.

Schlesselman J. Case-control Studies: Design, Conduct, Analysis. New York: Oxford University Press, 1982.

Schwarz N, Sudman S. Autobiographical Memory and the Validity of Retrospective Reports. New York: Springer Verlag, 1994.

Schwarz N, Sudman S. Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. 1 edn.

San Francisco: Jossey-Bass Publishers, 1996.

Shear MK, Brown TA, Barlow DH, Money R, Sholomskas DE, Woods SW, Gorman JM, Papp LA. Multicenter collaborative panic disorder severity scale. Am J Psychiatry 1997; 154: 1571–5.

Sirken MG, Herrmann DJ, Schechter S, Schwarz N, Tanur JM and Tourangeau R. Cognition and Survey Research. New York: John Wiley & Sons, 1999.

Skelton JA, Croyle RT. Mental Representation in Health and Illness. New York: Springer-Verlag, 1991.

Spitzer RL. Diagnosis and need for treatment are not the same. Arch Gen Psychiatry 1998; 55: 120.

Stewart A, Greenfield S, Hays RD. Functional status and well-being of patients with chronic conditions. JAMA 1989; 262: 907–13.

Straus MA, Hamby SL, Finkelhor D, Moore DW, Runyan D. Identification of child maltreatment with the Parent-Child Conflict Tactics Scales: development and psychometric data for a national sample of American parents. Child Abuse Negl 1998; 22: 249–70.

Substance Abuse and Mental Health Services Administration. Final notice establishing definitions for (1) children with a serious emotional disturbance, and (2) adults with a serious mental illness. Fed Regist 1993; 58: 29422–5.

Sudman S, Bradburn N, Schwarz N. Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass, 1996.

Sullivan MD, LaCroix AZ, Russo JE, Walker EA. Depression and self-reported physical health in patients with coronary disease: mediating and moderating factors. Psychosom Med 2001; 63: 248–56.

Tanur JM. Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation, 1992.

Tsang HW, Tam PK, Chan F, Chang WM. Sources of burdens on families of individuals with mental illness. Int J Rehabil Res 2003; 26: 123–30.

Turner C, Martin E. Surveying Subjective Phenomena. New York: Russell Sage Foundation, 1985.

Tweed DL. Depression-related impairment: estimating concurrent and lingering effects. Psychol Med 1993; 23: 373–86.

Üstün TB, Chatterji S and Rehm J. Limitations of diagnostic paradigm: it doesn't explain 'need'. Arch Gen Psychiatry 1998; 55: 1145–8.

van den Bos GA. The burden of chronic disease in terms of disability, use of health care and healthy life expectancies. Eur J Public Health 1995; 5: 29–34.

Verbrugge LM and Patrick DL. Seven chronic conditions: their impact on US adults' activity levels and use of medical services. Am J Public Health 1995; 85: 173–82.

Vinokur A, Oksenberg L and Cannell CF. Effects of feedback and reinforcement on the report of health information. In CF Cannell, L Oksenberg and JM Converse eds. Experiments in Interviewing Techniques. Ann

Arbor MI: Survey Research Center, University of Michigan, 1979.

Wells KB, Stewart A, Hays RD, Burnam M, Rogers W, Daniels M, Berry S, Greenfield S and Ware J. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. JAMA 1989; 262: 914–19.

WHO International Consortium in Psychiatric Epidemiology. Cross-national comparisons of the prevalences and correlates of mental disorders. Bull World Health Organ 2000; 78: 413–26.

Wittchen H-U. Reliability and validity studies of the WHO Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 1994; 28: 57–84.

Wittchen H-U, Üstün TB and Kessler RC. Diagnosing mental disorders in the community. A difference that matters. Psychol Med 1999; 29: 1021–7.

World Health Organization. Composite International Diagnostic Interview. Geneva, Switzerland: World Health Organization, 1990.

World Health Organization. WHO Disablements Assessment Schedule II (WHO-DAS II). Geneva, Switzerland: World Health Organization, 1998.

World Health Organization. International Classification of Functioning, Disability and Health. Geneva, Switzerland: World Health Organization, 2001.

Zeiss A, Lewinsohn P. Enduring deficits after remission of depression: a test of the 'scar' hypotheses. Behav Res Ther 1988; 26: 151–8.

Correspondence: RC Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA, USA 02115.
Telephone (+1) 617-432-3587.
Fax (+1) 617-432-3588.
Email: kessler@hcp.med.harvard.edu.

Also:

T Bedirhan Üstün, Global Programme on Evidence for Health Policy, World Health Organization, Avenue Appia 20, 1211 Geneva 27, Switzerland.
Telephone (+41) 22-791-3609.
Fax (+41) 22-791-4885.
Email: ustunb@who.ch.

Appendix: examples of WMH-CIDI modifications

Question comprehension

Breaking down complex questions into less complex sub-questions

A good example of a complex question is the stem question for dysphoria in the depression section. This is a very

important question because failure to endorse it or a parallel question about anhedonia leads to the respondent being skipped out of further questions about depression. The question in the standard CIDI is as follows: 'In your lifetime, have you ever had two weeks or longer when nearly every day you felt sad, empty, or depressed for most of the day?' This is an exceedingly complex question, as it asks about a cluster of emotions (sad, empty, or depressed) over a duration of time (2 weeks or longer), which itself can have internal variation in duration both across days (nearly every day) and within days (most of the day). Debriefing studies show that most respondents miss at least one of these core components. Therefore, we decompose the question in the WMH-CIDI. We begin by asking about a shorter period of time and omit mention of between-day variation in duration: 'Have you ever in your life had a period lasting several days or longer when most of the day you felt sad, empty, or depressed?' Positive responses are then followed by a duration question: 'Did you ever have a period of this sort that lasted <u>most of the day, nearly every day,</u> for <u>two weeks</u> or longer?' (The underlines are an indication to interviewers to emphasize these words.) This question is then followed by a within-day duration question: 'Think of times lasting two weeks or longer when these problems with your mood were more severe and frequent. During those times, did these feelings usually last less than one hour, between one and three hours, between three and five hours, or more than five hours?' The reader might think that this last question is unnecessary, as the question before it asked about dysphoria lasting 'most of the day'. However, pilot studies of the WMH-CIDI showed clearly that this aspect of the question was the least likely to be heard by respondents, leading to quite a few respondents who endorsed this question reporting in a follow-up question that their low mood lasted only for an hour or two. Based on this result, the follow-up question about within-day duration was retained in the final WMH-CIDI. A separate follow-up question about between-day duration (whether the dysphoria lasted every day, nearly ever day, most days, half the days, or less than half the days over the 2 weeks) was found not to be necessary, as all pilot respondents responded 'every' or 'nearly every day'.

Clarifying vaguely defined terms

A good example of a vaguely defined term is the single standard dichotomous CIDI question about role impairment that is used in every diagnostic section of the CIDI: 'Did (SYNDROME) interfere with your life and activities a lot?' This is a critical question, as it is used to operationalize the impairment component of the DSM-IV requirement that a syndrome cause clinically significant distress or impairment to qualify as a disorder. Yet both the word 'interfere' and the words 'a lot' are ambiguous.

Furthermore, the fact that the question is dichotomous, and that it requires an intensity qualifier to be answered positively (it is not enough for the syndrome to interfere; it must interfere a lot), creates a source of confusion. This was clearly visible in pilot studies carried out by Kessler et al. (2000), who found that respondents who endorsed the standard CIDI question often responded 'some' or 'a little' to a follow-up question that asked 'How much did it interfere with these activities – would you say a lot, some, or only a little?'

It might seem strange that a respondent who has just responded 'yes' to a question about whether there was 'a lot' of interference would characterize the interference as less intense in the very next question, but the fact that this frequently occurs illustrates that survey respondents often fail to listen carefully to secondary clauses in dichotomously worded questions. Recognizing this problem, the WMH-CIDI converts the dichotomous yes-no question about a lot of interference into a dimensional question that focuses the respondent's attention on the intensity of interference by asking 'How much did [SYNDROME] interfere . . . – not at all, a little, some, a lot, or extremely?' A small split ballot experiment in a pilot study carried out by Kessler et al. (2000) found that this wording change resulted in a significantly lower proportion of people reporting that the interference was severe enough to be characterized as 'a lot' than in the standard dichotomous CIDI question.

In addition to turning the dichotomous response scale into a polychotomous scale, the WMH-CIDI also expands the standard CIDI interference question to make sure respondents broadly review all major areas in their life before answering. The fully modified question is: 'How much did [SYNDROME] interfere with either your work, your social life, or your personal relationships – not at all, a little, some, a lot, or extremely?' In order to clarify the intensity level implied by the various response categories, a follow-up question is then asked in the WMH-CIDI to all respondents who report any interference: 'How often were you unable to carry out your daily activities because of [SYNDROME] – often, sometimes, rarely, or never?'

Clarifying questions about odd experiences that could be normalized

The standard version of CIDI introduces the psychosis questions with the statement: 'Now I want to ask you about some ideas you might have had about other people.' The questioning then begins by asking about delusions rather than about hallucinations. The first question is: 'Have you ever believed people were spying on you?' This is followed by questions about believing that people are spying on you and talking about you behind your back. Each time a positive response is given, the interviewer

asks: 'How did you know this was happening?' An open-ended response is recorded and rated for plausibility. A great many people answer these questions positively, the vast majority of whom give plausible answers. This is not surprising, as the experiences asked about are all quite common. The first genuinely odd experience isn't asked about until the fourth question in the series: 'Have you ever believed that you were being secretly tested or experimented on?' This is followed by two remaining questions about delusions that could be interpreted in plausible terms: whether someone ever 'plotted against you' or 'tried to hurt you or poison you'; whether the respondent ever thought that 'someone you never met was in love with you'; and a question about whether the respondent ever was convinced that his or her spouse or partner was being unfaithful even though they said this was not true.

These are followed by questions about more bizarre delusions, such as mind reading (for example, 'Have you ever believed that someone was reading your mind?') thought control (for example, 'Have you ever been convinced that you were under the control of some power or force, so that your actions and thoughts were not your own?'), and being able to receive special messages through the mass media. Finally, the symptom assessment finishes with questions about visual, auditory, olfactory, and tactile hallucinations. Most of these questions are purposefully worded in a way that is designed to normalize them. For example, the auditory hallucinations question asks: 'Have you more than once heard things other people couldn't hear, such as a voice.' This normalized phrasing leads to confusion, as a great many respondents in general population samples respond positively and then, in response to the follow-up question 'How do you explain hearing things other people couldn't hear?' respond by saying 'I have very good hearing.'

The philosophy behind this approach to assessing psychosis is that psychotics will be more willing to admit their symptoms if these symptoms are normalized. A great difficulty with this approach is that it generates an enormous number of false positives. This substantially complicates the process of screening for psychosis and also introduces the strong possibility of errors in classifying false positives as cases based on misleading open-ended responses. In addition, a question can be raised whether true psychotics, particularly those with paranoid tendencies, will be motivated positively or negatively to respond positively to normalized questions that they might perceive as trying to trick them into reporting their clearly abnormal experiences. The philosophy behind the WMH-CIDI approach is the opposite: to make it clear to respondents that we are asking about odd experiences; to motivate reporting with an introduction that validates the experiences and points to the importance of learning more about them; and beginning the questioning with hallucinations

rather than delusions in order to reinforce the introductory remarks about the questions being about odd behaviours.

The WMH-CIDI introduction is as follows: 'The next questions are about unusual things, like seeing visions or hearing voices. We believe that these things may be quite common, but we don't know for sure because previous research has not done a good job of asking about them. So please take your time and think carefully before answering.' One can see in this introduction a number of the methodological refinements discussed in the body of the paper: clarification of the nature of the questions; validation of the experiences; motivation for honest reporting; and facilitation of serious memory search by legitimating the respondent not answering immediately in order to think. The questions themselves are worded in such a way as to avoid confusion. Compare, for example, the WMH-CIDI question about auditory hallucinations with the standard CIDI question about the same symptom: 'The next question is about hearing voices that other people could not hear. I don't mean having good hearing, but rather hearing things that other people said did not exist, like strange voices coming from inside your head talking to you or about you, or voices coming out of the air when there was no one around. Did you ever hear voices in this way?'

Contextual misunderstanding

Contextual misunderstanding is a type of misunderstanding that derives from the position of the question in the flow of the interview. An example was given there about confusion in a question regarding the duration of phobic fears. The standard question order in the CIDI was modified to correct this problem. A number of similar small but important cases of a related sort were found and corrected in developing the WMH-CIDI. One additional example: in asking about number of lifetime manic episodes, the standard CIDI comes directly after a question about the duration of the respondent's longest lifetime manic episode. The number-of-episodes question then follows, asking about the number of 'these' lifetime episodes. Debriefing shows that the word 'these' in the context of this placement leads a number of respondents incorrectly to believe that the interviewer is asking about episodes that were equally as long as the longest lifetime episode, resulting in an underestimate of the number of lifetime manic episodes. A person who had, say, 30 manic episodes, only four of which went on as long as the longest episode, might respond 'three' rather than '30' to this question about number of episodes. This problem was corrected in the WMH-CIDI by a change in question placement.

Task comprehension

Survey respondents are often unaware that interviewers want them to engage in active memory search in

answering complex questions. Indeed, it often occurs that investigators who write survey questions without cognitive debriefing pilot studies themselves fail to appreciate the cognitive complexity involved in answering some survey questions. This is all the more true for interviewers, who, in the absence of special training in the use of feedback methods aimed at stimulating thoughtful responding, will be more concerned with the accuracy of recording answers than in the accuracy of the answers being given. The WMH-CIDI includes clarifications in a number of places to make this task clear to respondents. Earlier, the paper discussed one of the most important places where this is done, in the life review section of the interview. See this section of the paper for the example.

There are numerous other places in the instrument where the same principle is used to remind the respondent to think carefully. One of the most consistent of these is in the important question about age of disorder onset, which appears in each diagnostic section of the interview. The standard CIDI question asks 'How old were you the first time you [HAD THE SYNDROME].' We know from debriefing interviews that many respondents have a very difficult time remembering their age of onset, especially for accretion disorders. Therefore, it is important both to make it clear that we are looking for a serious memory search and to deal with the possibility that we may be asking for more than the respondent knows. As described in the body of the paper, we deal with the first of these issues in the WMH-CIDI by asking what is known in the survey methodology literature as a 'prequest', a question designed to create a clarifying context for the substantive question that follows it: 'Can you remember your exact age the very first time you [HAD THE SYNDROME]?' Even though a great many respondents answer 'no', this question is extremely important in making the task clear to respondents. We are interested in a precise answer, which means that serious memory search will be required.

It is also noteworthy that we try to be equally clear when we want estimates rather than precise information. This is important because it is not possible to motivate respondents to engage in active memory search throughout an entire long interview. Instead, we need to pick and choose the especially important questions where extra effort is thought to be needed as well as to be clear both to ourselves and to respondents when we are willing to settle for approximations. So, for example, questions concerning lifetime course of illness (for example, 'About how many different years in your life did you have [SYNDROME]?') explicitly ask for approximations by using the word 'about'. In addition, interviewers are trained both to accept approximations as answers to these questions and to probe for rough estimates if respondents say they are unable to provide exact responses. In a similar way, we use structured response categories with prespecified ranges rather than open-ended responses in cases where we ask estimation questions that will be difficult to answer and we recognize the inability to obtain accurate fine-grained responses from most respondents.

Motivation

Motivational instructions and commitment questions

We developed a statement at the beginning of the life review section to emphasize the importance of careful response in order to encourage complete and accurate answers. As noted earlier, a commitment question is used in conjunction with the motivational instructions to make sure that the respondent is aware of and acknowledges the importance of responding carefully. The statement and commitment question are as follows: '[READ SLOWLY] The next questions are about health problems you may have had at any time in your life. It is important for the research that you think carefully before answering. Are you willing to do this?' In cases where the respondent does not answer 'yes' to the commitment question, the interviewer is instructed to offer to reschedule the interview for a time when the respondent is more able to give serious thought to the questions. If respondents persist in saying that they are unwilling to think carefully before answering, the interviewer is instructed to terminate the interview and to code the respondent as a refusal.

Contingent reinforcement

Contingent reinforcement is an interviewer feedback strategy rather than a question-wording strategy. The interviewer training manual for the WMH-CIDI (which is available only to researchers who participate in the official WMH-CIDI training described earlier in the paper) focuses considerable attention on the use of this directive feedback strategy. A variety of probes are provided to interviewers to reinforce apparent effort in providing thoughtful answers. These are very simple feedback responses, like 'thanks' and 'that's useful', which are administered whether the response is positive or negative, so long as effort appears to be invested in providing a thoughtful answer. Responses that might be thought to imply a value judgment, like 'good' or 'excellent', are not used.

A variety of probes are also provided to give negative feedback for apparently superficial responses. Sometimes these take the form of a follow-up question to a seemingly superficial negative response (for example, 'are you sure there was nothing?'). At other times they take the form of an injunction (for example, 'please take your time and think carefully'). At still other times the probe can take the form of an observation followed by a request (for example, 'You answered that one awfully quickly. Could

you take a minute to think hard and make sure you didn't forget anything?'). Our experience is that the long versions of the negative probes are most useful early in the interview, when respondents who give superficial answers are being trained to be more thoughtful. Periodic short positive feedback probes, with the rare use of a short negative probe at the first signs of the recurrence of superficial responding, generally are used throughout the remainder of the interview.

The ability to answer accurately

We use two broad strategies to deal with the realization that autobiographical memory has limits. The main one is to accept these limits and revise our questions so as to settle for less detail than we would ideally like to have. The other is to push the limits of memory by decomposing questions in ways that mimic successful memory search strategies and bound uncertainty. We already gave examples earlier in the appendix of the first of these two strategies. As illustrated in these examples, we always signal to respondents when we want approximations rather than exact responses. This is done either by building into the question an explicit reference to wanting an approximation or by providing structured response categories in rough groupings that indicate the level of approximation we are looking for in responses. Perhaps the best example of this approach is the question series described earlier in the appendix about age of disorder onset. As noted in that earlier description, we begin this series with the prequest: 'Can you remember your <u>exact age</u> the <u>very first time</u> you [HAD THE SYNDROME]?' Respondents who answer 'yes' are dealt with easily by asking them to report this exact age. Respondents who say 'no', in comparison, are asked for an estimate: '<u>About</u> how old were you the first time?'

This same question series also illustrates the use of the second strategy used to deal with the limits of autobiographical memory: decomposing questions and bounding uncertainty for respondents who vary in accuracy of recall. As information about age of onset is of great importance for a variety of substantive research questions, special effort was invested in pushing for the limits of memory among respondents who reported that they could not recall the exact age when the syndrome first began. This is done using special probes to ask respondents to go back-

wards in time sequentially. This requires flexibility on the part of interviewers, as a good many respondents who answer 'no' to the question about exact recall volunteer, in conjunction with that answer, that the syndrome has been going on 'my whole life' or 'as long as I can remember'.

This kind of inexact information is useful because it implies a very early age of onset. To confirm this, though, such responses are probed for clarification of an upper bound by asking 'Was it before you first started school?' If the respondent answers 'yes', the question sequence ends, with a fairly narrow bounding of uncertainty in an early age range. If, however, the respondent does not answer 'yes' to this first probe (for example, if the respondent says 'I can't really remember' or 'I don't think so' or something similar), interviewers then move up the age range incrementally until they find an interval of time at which the respondent feels secure in saying that the syndrome was definitely in existence as of that point in the life course (for example, 'Was it before you were a teenager?' If not, 'Was it before you turned 20?' and so forth). Our experience in pilot studies, which has been confirmed in WMH surveys, is that most uncertainty among respondents who begin by saying that their inability to provide an exact age is due to the syndrome going on 'as long as I can remember' can be bounded before the teenage years.

Among respondents who do not volunteer a comment to the effect that the syndrome has been going on 'as long as I can remember' in response to the exact age question, in comparison, the interviewer begins by probing for such a response ('Did it go on as long as you can remember?'). If so, then probing proceeds as if this response was volunteered. If respondents say that the syndrome did not go on as long as they can remember, in comparison, debriefing studies show that recall is quite certain for a fairly specific time in the life course, but not for a particular age. Therefore, the same series of probes moving up the age range is used as described in the last paragraph, with a final request for a best estimate of the onset age once an upper bound age range is reached. Responses that are given in ranges (for example, 'It started during my teens, but I can't remember the exact age') are recorded initially as ranges and then analysed as the upper end of the range in order to provide a consistently conservative lower bound estimate.