

EDITORIAL

Diagnosing mental disorders in the community.
A difference that matters?¹

Brugha and his colleagues in this issue raise important questions about the validity of standardized diagnostic interviews of mental disorders, such as the Composite International Diagnostic Interview (CIDI) (WHO, 1990). Although their concerns refer predominantly to the use of such instruments in epidemiological research, the authors' conclusions also have significant implications for diagnostic assessments in clinical practice and research. We agree with Brugha *et al.* that the inflexible approach to questioning used in standardized interviews can lead to an increased risk of invalidity with regard to some diagnoses. We also agree that the use of more semi-structured clinical questions has the potential to address this problem. However, we disagree with Brugha *et al.* in several other respects.

First, we disagree with the authors' initial exclusive emphasis on diagnosis with regard to need assessment and consequences for the allocation of service resources. It is becoming increasingly clear that knowledge about diagnosis does not, in itself, whether assessed by clinical or non-clinical diagnostic interviews, provide sufficient information we need for policy purposes and the determination of societal costs, or to judge clinical management guidelines and treatment needs (Regier *et al.* 1998). Additional, preferably dimensional, data on associated disabilities and distress as well as a focused need evaluation for those psychosocial, psychological and drug interventions that characterize modern treatment strategies are also important. It also has become evident that a great many people in the general population carry more than one diagnosis. This 'co-morbidity' complicates further such simple equation of diagnosis prevalence with need assessment and policy decisions. Secondly, we disagree with the conclusion of Brugha *et al.* that the use of a semi-structured clinical interview, like the most current version of the Structured Clinical Assessment for Neuropsychiatry (SCAN), whether in the hands of clinical or non-clinical interviewers, is most closely approximating the 'clinical gold standard' and is the most feasible way to correct the problem of disagreement between semi-structured clinical diagnostic interviews and standardized diagnostic interviews. We believe that the practical reliability and validity problems associated with using such a clinical interviewing approach especially in large-scale community surveys as well as in cross-national research more than cancel out any theoretical advantage this approach might have in clarifying meaning. Thirdly, we disagree with the suggestion of Brugha *et al.* that the problem of validity is inherent in standardized non-clinician interviews. Indeed, as detailed below, there is no evidence that across all diagnoses clinical semi-structured interviews reveal more promising psychometric properties than standardized interviews. Also methodological research shows quite clearly that a substantial number of potential validity problems in standardized interviews can be overcome.

Based on these considerations, we believe that the best way forward is: (a) to retain the standardized diagnostic interview as the method of choice in community epidemiological surveys as well as clinical epidemiology; (b) to include both categorical assessments of diagnosis and dimensional assessments of disability and distress; and (c) to refine the standardized assessments of both categories and dimensions by improving question and response category wording and by using well established clinician interviewing strategies (i.e. open-ended questioning with optional

¹ Address for correspondence: Professor Hans-Ulrich Wittchen, Clinical Psychology and Epidemiology, Max Planck Institut für Psychiatrie, Kraepelinstrasse 10, 80804 Munich, Germany.

structured probes along with interviewed-based response ratings) to assess a limited number of especially difficult diagnostic and clinical constructs. We also believe that appropriate semi-structured clinical reappraisal interviews administered to a probability subsample of survey respondents should be a routine part of epidemiological survey methodology. However, we differ from the approach taken by Brugha *et al.* in recognizing that the results of clinical reappraisal interviews should be interpreted with a realization that clinical interviews, especially those administered in community settings, might have serious problems of both reliability and validity as well.

Before elaborating this further a few more general remarks seem to be necessary in challenging Brugha *et al.*'s emphasis on 'clinical judgement' as key prerequisite of diagnostic assessment.

THE QUEST FOR CONSISTENCY IN DIAGNOSTIC ASSESSMENT

A major ongoing effort to develop objective and reliable diagnostic classifications of mental disorders began with the DSM-III and has continued in successive revisions of both the ICD and DSM systems. Among the key motivations for this effort were the repeated findings that diagnostic decisions in mental health research and practice suffer from low inter-rater agreement (Matarazzo, 1983; Robins & Barret, 1989; Berger *et al.* 1993) and the realization that acceptance of mental health treatment as a science rather than an art form requires the existence of explicit and more objective diagnostic criteria. Based on these considerations, successive refinements to diagnostic classification in the ICD and DSM systems over the past two decades have attempted to emphasize more and more explicit symptom descriptions and criteria, thus reducing the previously central, however only vaguely defined role of 'clinical judgement' in psychopathology and diagnostic assessment. As a result of this process ICD-10 and DSM-IV not only include many more pages, but also more diagnoses, new diagnostic principles (i.e. multiple diagnoses) and more subtle diagnostic distinctions of previously fairly broad nosological concepts (i.e. neuroses, addictions), that imply in various aspects a departure from several traditional, yet insufficiently proved and questionable nosological assumptions of mental illness. Therefore, it is not surprising to see especially in many European countries with their rich traditions of 'schools' of psychopathology, that most of the changes are up to now only partly reflected in standard textbooks, training, research and practice and certainly they are not uniformly appreciated and adopted. Against this background it seems to be fair to state that there is no uniformly accepted procedural and content definition of 'clinical judgement', that remains a fairly vague construct, difficult to grasp scientifically.

Efforts to improve structured psychopathological assessment schedules and interviews have been closely linked to this ongoing process (Robins, 1985). In fact, the development of semi-structured clinical diagnostic interviews and the introduction of its much more radical next step, namely the fully standardized diagnostic interviews, have only been possible by the existence of explicit descriptive diagnostic criteria and the specification of diagnostic algorithms for hundreds of specified diagnoses of mental disorders (Robins, 1987). Whereas before the introduction of DSM-III such interviews could at best offer 'ICD compatible classes', the new generation of diagnostic interviews do offer a highly objective computation of specific diagnoses. It is now routine for research diagnostic interviews to require systematic and comprehensive review of all diagnostic criteria, explicit question wording to initiate more or less strictly defined probing of specific symptoms, and consensus standards to define thresholds for dichotomous symptom and disorder classifications of open-ended responses. With careful training and close monitoring to prevent inter-rater drift, it is now possible to obtain good inter-rater agreement on diagnostic ratings of patients using currently available semi-structured research diagnostic interviews (Williams *et al.* 1992; Brugha *et al.* 1999).

For purposes of epidemiological research, though, it is important to recognize that agreement is lower for borderline cases of the sort that are likely to be found in community samples rather than in clinic samples and even lower when it comes to the assessment of past and not current (acute) conditions. It is also important to recognize that interviewer drift can easily occur and that maintenance of high inter-rater agreement requires careful, consistent, and ongoing monitoring of

interviewer performance of a sort that, as a practical matter, cannot occur in large-scale community epidemiological surveys carried out by a widely dispersed field force scatters over an entire country. Even more problems are typically experienced, when clinical interview standards have to be established across countries, due to the above mentioned variability of psychiatric schools of thoughts with regard to implicit and explicit nosological concepts that influence the content of 'clinical judgement'. As a result, while it is appropriate to think of using clinical reappraisal interviews administered by a small, carefully selected, and closely supervised field force as a validation standard in smaller regional or national community epidemiological surveys, the use of semi-structured clinical interviews as the central approach to carrying out such surveys on a larger scale is likely to create more problems than it solves.

It was in recognition of these facts that the former US-Alcohol, Drug and Mental Health Administration/WHO Task Force together with many experts around the world sponsored the development of the standardized Composite International Diagnostic Interview (CIDI) primarily for use by non-clinicians in parallel with the semi-structured clinician administered Schedule for Clinical Assessment in Neuropsychiatry (SCAN) (Wing *et al.* 1998). The CIDI was designed primarily to provide a reliable assessment of symptoms and diagnoses according to ICD-Diagnostic Criteria for Research and DSM-III-R/IV in large-scale community surveys and other settings in which it is impractical to use carefully trained and closely supervised clinical raters. The SCAN was designed to provide a much more comprehensive (in terms of more subtle ratings and a more comprehensive coverage of psychopathological symptoms) assessment of psychopathology in clinical settings. It was not the intention of WHO that the CIDI be considered an inferior version of the SCAN, but that the two instruments be complementary and useful in different settings. Consistent with this perspective, we reject the notion of Brugha *et al.* that the SCAN should be considered the validation standard for the CIDI. As discussed below, there are clearly cases where the CIDI would be expected to perform better than the SCAN.

THE PROBLEM OF VALIDITY IN COMMUNITY SURVEYS

We agree with Brugha and colleagues, that diagnostic interviews are imperfect and consequently might pose problems of validity in community surveys as well. But is there any empirical evidence that clinical diagnostic interviews, and the SCAN specifically, due to its emphasis on 'clinical judgement' fair better than the standardized interviews? Brugha *et al.* fail to provide such evidence and also fail to acknowledge quite a number of inter-rater and test-retest reliability studies as well as several clinical reappraisal studies conducted with standardized interviews (review by Wittchen, 1994). Disregarding the fact, that there is a much more comprehensive and at least five-fold higher psychometric database for standardized interviews, these studies clearly do not demonstrate any superiority of the SCAN approach over the CIDI approach in terms of reliability (rather the other way around). We agree with Brugha and colleagues that the database of clinical validity studies is less impressive as well as their conclusion that both types of instruments have similar limitations in such 'that a high proportion of important and potentially treatable disorder ... will be incorrectly identified'. But at least CIDI methodology studies have provided considerable evidence for which diagnoses this type of instrument performs well and such studies also identified specific limitations opening up strategies of further improvement.

Survey methodologists have long been concerned with the problem of validity of self-report measures including symptoms of mental state functioning (Bradburn *et al.* 1979; Belson, 1981; Cannell, 1985). Research on this topic has made major strides forward in recent years based on collaborations between survey methodologists and cognitive scientists using insights from cognitive research on basic processes of understanding and motivation (Jabine *et al.* 1984; Skelton & Croyle, 1991; Schwarz & Sudman, 1993). We now know a great deal, based on these new cognitive survey studies, both about the reasons for invalidity in surveys and about effective strategies to minimize invalidity (Tanur, 1992; Schwarz & Sudman, 1996; Sudman *et al.* 1996; Knäuper *et al.* 1999).

Brugha *et al.* focused in their editorial on one of the four main reasons for survey invalidity

documented in this methodological literature: respondent misunderstanding of questions and response categories. They correctly point out that misunderstanding can be a problem in fully structured diagnostic interviews and that this problem might possibly be reduced by using a more open-ended interviewing style that clarifies the meanings of questions and of response categories. However, they displayed no awareness of the literature on survey methodology and consequently made two critical implicit assumptions that this literature clearly shows to be false: that clinical interviewing is the *only way* to correct the problem of misunderstanding in surveys; and that misunderstanding is the *main reason* for invalidity in surveys. In the next two sections we will briefly elaborate on these two misunderstandings.

STRATEGIES TO CORRECT THE PROBLEM OF QUESTION MISUNDERSTANDING

Think aloud experiments (Anders *et al.* 1993), cognitive debriefing interviews (Groves *et al.* 1992) and other cognitive survey techniques (Fowler & Cannel, 1996) have documented that a number of strategies are effective in detecting and correcting the problem of question misunderstanding in standardized interviews. In addition to the obvious strategy of using words and phrases that are as unambiguous as possible, the tools available to the question designer include clarifying introductions, exemplar lists, visual aids, and vignette-type response categories (many of which have become integral components of the most recent revisions of the CIDI) (Wittchen *et al.* 1998*a*). In situations where the meaning of a question (and the underlying clinical construct) is difficult to communicate to respondents, it is possible to train interviewers to select from a set of optional probes to elicit open-ended responses that can subsequently be coded by clinically trained coders after the interview is completed, without introducing sources of variance that are difficult to control for.

As described elsewhere (Kessler *et al.* 1998), we evaluated Version 1.1 of the World Health Organization's (WHO) Composite International Diagnostic Interview (CIDI) (WHO, 1990) using cognitive survey techniques and uncovered a number of important areas of systematic question misunderstanding that we were able to correct in versions of the CIDI used in surveys carried out in the US (Kessler *et al.* 1994) and Germany (Wittchen *et al.* 1998*a,b,c*). More recent methodological studies carried out in preparation for the WMH2000 surveys have led to further CIDI refinements aimed at minimizing the effects of question misunderstanding. Preliminary comparisons of this revised version of the CIDI with blind clinical reappraisal interviews in community samples show much higher rates of agreement than those found by Brugha *et al.* in their evaluation of the CIS-R.

Results such as these argue that Brugha and his colleagues were too quick to reject standardized interviews as inherently flawed and to conclude that 'flexible cross-questioning and rating judgement are critically important' for the valid assessment of mental disorders in community surveys of mental disorders. It is certainly true that clarity of meaning is necessary to obtain meaningful responses. But the use of semi-structured clinical interviews is only one of a number of strategies that are capable of clarifying meaning. This conclusion is consistent not only with our methodological studies of the CIDI, but also with recent work in medical informatics that has successfully developed fully-structured computerized self-administered screening versions of well-known clinical severity scales for use in patient tracking. Fully structured self-administered versions of the Hamilton Rating Scale of Depression (Reynolds & Kobak, 1995) and the Liebowitz Social Anxiety Scale, for example, have both been shown to have extremely high correlations with clinician-administered ratings.

BEYOND QUESTION MISUNDERSTANDING: OTHER SOURCES OF INVALIDITY

We noted at the beginning of this Editorial that question understanding is only one of the sources of invalidity studied by survey methodologists. Others include understanding the task (i.e. understanding that thoughtful and honest responding is expected), being willing to carry out the

task (i.e. being motivated to provide thoughtful and honest responses), and being able to carry out the task (i.e. dealing with questions that exceed the limits of memory). Although Brugha and his colleagues make no mention of them in their critique, these problems are generally considered to be more important than question misunderstanding as sources of invalidity in community surveys (Biemer *et al.* 1991). This is especially likely to be true in surveys of mental illness, which deal with potentially embarrassing material that respondents may be reluctant to discuss.

Survey methodologists have developed a number of strategies to deal with these problems (Sudman *et al.* 1996). Semi-structured clinical interviews have no particular advantage over fully structured lay interviews in implementing these strategies. Indeed, in the case of the problem of willingness to provide thoughtful and honest responses, there is good reason to believe that lay interviews have an advantage over clinician-administered interviews. This advantage is associated with the fact that methodological studies consistently find that honesty of response to questions about embarrassing and painful experiences increases as the psychological distance increases between the researcher and the respondent (Groves, 1989; Turner *et al.* 1992). This means that the engaging and conversational style of clinical interviews can be a disadvantage compared to the more mechanical and distant style of standardized interviews. A special advantage of standardized assessment is that it allows especially embarrassing questions to be self-administered, an approach that has been shown to have a dramatic effect on reports about embarrassing behaviours and experiences (Turner *et al.* 1998).

THE VALIDITY OF CLINICAL VALIDATION STUDIES IN COMMUNITY SAMPLES

This last issue raises questions about the validity of clinical validation studies such as that reported by Brugha and his colleagues in this issue. If community respondents are less willing to admit embarrassing and painful experiences in semi-structured clinical validation interviews than in standardized lay interviews, then the many inconsistencies classified by Brugha *et al.* as 'false positive' lay interview diagnoses might, in fact, be false negative clinical interview diagnoses.

It is important in this regard to recognize that the norm of disclosure embodied in the doctor-patient relationship, which would be expected to reduce embarrassment in a clinical setting (Browne & Freeling, 1967), does not operate in a community interview situation. A very telling illustration of this fact was reported by Reissman (1977), who embedded a field experiment study in a community survey where all interviews were carried out by psychiatric residents. The experimental manipulation involved the interviewers randomly informing respondents at the beginning of the survey that they were 'interviewers' or 'psychiatrists'. Identical questions were asked in the two experimental conditions. Reissman found that significantly lower prevalences of disorder were reported by low income, male, and minority respondents who were informed that their interviewers were psychiatrists. Reissman interpreted this finding as due to psychiatrists being perceived as social control agents (e.g. people who can sign a court order and have one's children taken away) in some segments of society, leading respondents in these social strata to hide, rather than to disclose, information about mental illness. In the face of evidence such as this, it is very difficult to maintain the view espoused by Brugha *et al.* that a clinical reinterview is the gold standard against which standardized diagnostic interviews in community surveys should be judged.

There are at least two other reasons to call into question the validity of conventional clinical validation studies in community surveys. One is acknowledged in passing by Brugha *et al.* in their editorial: that the reliability of the SCAN has not been established in community samples. The other is that respondent fatigue leads to lower prevalence estimates in community survey reinterviews than initial interviews (Helzer *et al.* 1981). Importantly, this pattern of lower prevalences in reinterviews has been found even when the same semi-structured clinical interview was administered twice (Bromet *et al.* 1986). Given the low short-term test-retest reliability of semi-structured clinical interviews and the artificial reduction in estimated prevalences in reinterviews due to fatigue, we would expect low concordance between structured interviews and clinical reinterviews and higher

prevalences in the lay interviews than clinical interviews even if the lay interview was perfectly reliable and valid.

WHAT IS THE BEST WAY FORWARD?

Based on the positive results of a pilot study in a patient sample, Brugha *et al.* suggested that the best way forward in dealing with the problem of validity in community surveys of mental illness might be to train lay interviewers to administer the SCAN. This is a strategy that has been used for many years by Brown & Harris (1978) in their studies of life events and depression. It is consequently not surprising that Brugha documented good agreement between diagnoses based on lay-administered and clinician-administered SCAN interviews in his pilot study. This is especially true in light of the fact that the sample in this pilot study was made up largely of in-patients and that the interviewers consisted of a small, highly trained (320 hours of training per interviewer), group carefully selected from the large field staff of the Office for National Statistics in London.

Despite the apparent success of this pilot study, we strongly disagree with Brugha *et al.* that administration of the SCAN by trained lay interviewers is a sensible way to carry out large-scale community epidemiological surveys. As noted above, the problem of question misunderstanding can be resolved in carefully constructed standardized interviews. There is not a meaningful advantage of semi-structured clinical interviews in this regard. When we combine this with the realization that semi-structured clinical interviews have major disadvantages that are avoided in standardized interviews (i.e. lower inter-rater reliability and lower respondent willingness to admit embarrassing material), the superiority of standardized assessment becomes apparent.

Based on these considerations, we believe a much more appropriate way forward than the one proposed by Brugha *et al.* is to recognize that standardized interviews are necessary, to make use of advances in research on question wording to construct optimally valid questions in such interviews, and to appreciate that question wording is only one of the several problems that can compromise survey validity. This is not to say that clinical interviews are irrelevant. Far from it. We believe that semi-structured clinical reappraisal interviews are an important adjunct that should be a routine part of all community epidemiological surveys. Indeed, we are including a very extensive clinical reappraisal component in the WHO WMH2000 surveys. However, in doing this we are being sensitive to the fact that clinical interviews are imperfect and that we cannot simply accept clinical evaluations of community cases at face value without recognizing and adjusting for their low inter-rater reliability and potential validity problems.

H.-U. WITTCHEN, T. B. ÜSTÜN AND R. C. KESSLER

REFERENCES

- Anders Ericsson, K. & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (Revised Edition). MIT Press: Cambridge.
- Belson, W. A. (1981). *The Design and Understanding of Survey Questions*. Gower: Aldershot.
- Berger, M., Möller, H. J. & Wittchen, H.-U. (1993). *Psychiatrie als empirische Wissenschaft*. W. Zuckschwerdt Verlag: München.
- Biemer, P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A. & Sudman, S. (1991). *Measurement Errors in Surveys*. John Wiley & Sons, Inc.: New York.
- Bradburn, N., Sudman, S. & Associates. (1979). *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. Jossey-Bass: San Francisco.
- Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A. & Schulberg, H. C. (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry* 43, 435–440.
- Brown, G. E. & Harris, T. (1978). *The Social Origins of Depression: A Study of Psychiatric Disorder in Women*. Free Press: New York.
- Browne, K. & Freeling, P. (1967). *The Doctor–Patient Relationship*. E. & S. Livingstone: Edinburgh.
- Brugha, T. S., Bebbington, P. E., Jenkins, R., Meltzer, H., Taub, N. A., Janas, M. & Vernon, J. (1999). Cross validation of a general population survey diagnostic interview: a comparison of CIS-R with SCAN ICD-10 diagnostic categories. *Psychological Medicine* 29, 1029–1042.
- Cannell, C. F. (1985). Experiments in the improvement of response accuracy. In *Survey Interviewing: Theory and Techniques* (ed. T. W. Beed and R. J. Stimson), pp. 24–62. Allen & Unwin: Winchester, MA.
- Fowler, F. J. & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research* (ed. N. Schwarz and S. Sudman), pp. 15–36. Jossey-Bass Publishers: San Francisco.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. John Wiley & Sons: New York.
- Groves, R. M., Fultz, N. H. & Martin, E. (1992). Direct questioning about comprehension in a survey setting. In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys* (ed. J. M. Tanur), pp. 49–61. Russell Sage Foundation: New York.
- Helzer, J. E., Robins, L. N., Croughan, J. L. & Welner, A. (1981). Renard Diagnostic Interview: its reliability and procedural validity

- with physicians and lay interviewers. *Archives of General Psychiatry* **38**, 393–398.
- Jabine, T., Straf, M., Tanur, J. M. & Tourangeau, R. (1984). *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*. Report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology. National Academy Press: Washington, DC.
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U. & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry* **51**, 8–19.
- Kessler, R. C., Wittchen, H.-U., Abelson, J. M., McGonagle, K. A., Schwarz, N., Kendler, K. S., Knäuper, B. & Zhao, S. (1998). Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. *International Journal of Methods in Psychiatric Research* **7**, 33–55.
- Knäuper, B., Cannell, C. F., Schwarz, N., Bruce, M. L. & Kessler, R. C. (1999). Improving accuracy of major depression age of onset reports in the US National Comorbidity Survey. *International Journal of Methods in Psychiatric Research* **8**, 39–48.
- Matarazzo, J. D. (1983). The reliability of psychiatric and psychological diagnosis. *Clinical Psychology Review* **3**, 103–145.
- Reissman, C. K. (1977). Interviewer effects in psychiatric epidemiology: a study of medical and lay interviewers and their impact on reported symptoms. *American Journal of Public Health* **69**, 485–491.
- Regier, D. A., Kaelber, C. T., Rae, D. S., Farmer, M. E., Knäuper, B., Kessler, R. C. & Norquist, G. S. (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders: Implications for research and policy. *Archives of General Psychiatry* **55**, 109–115.
- Reynolds, W. M. & Kobak, K. A. (1995). *Hamilton Depression Inventory: A Self-report Version of the Hamilton Depression Rating Scale (HDRS)*. Psychological Assessment Resources, Inc.: Odessa, FL.
- Robins, L. N. (1985). Epidemiology: reflections on testing the validity of psychiatric interviews. *Archives of General Psychiatry* **42**, 918–924.
- Robins, L. N. (1987). The assessment of psychiatric diagnosis in epidemiological studies. In *Psychiatric Epidemiology. Section Five of Psychiatric Update: The APA Annual Review, Vol. 6* (pp. 589–606). American Psychiatric Press: Washington, DC.
- Robins, L. N. & Barrett, J. E. (1989). *The Validity of Psychiatric Diagnoses*. Raven Press: New York.
- Schwarz, N. & Sudman, S. (1993). *Autobiographical Memory and the Validity of Retrospective Reports*. Springer-Verlag: New York.
- Schwarz, N. & Sudman, S. (1996). *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. Jossey-Bass Publishers: San Francisco.
- Skelton, J. A. & Croyle, R. T. (1991). *Mental Representation in Health and Illness*. Springer-Verlag: New York.
- Sudman, S., Bradburn, N. M. & Schwarz, N. (1996). *Thinking about Answers: the Application of Cognitive Processes to Survey Methodology*. Jossey-Bass Publishers: San Francisco.
- Tanur, J. M. (1992). *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. Russell Sage Foundation: New York.
- Turner, C. F., Lessler, J. T. & Gfroerer, J. C. (1992). *Survey Measurement of Drug Use: Methodological Issues*. Government Printing Office: Washington, DC.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H. & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science* **280**, 867–873.
- Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L. & Davies, M. (1992). The structured clinical interview for DSM-III-R (SCID): multisite test–retest reliability. *Archives of General Psychiatry* **49**, 109–133.
- Wing, J. K., Sartorius, N. & Üstün, T. B. (1998). *Diagnosis and Clinical Measurement in Psychiatry. A Reference Manual for SCA/PSE-10*. Cambridge University Press: Cambridge.
- Wittchen, H.-U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI) – a critical review. *Journal of Psychiatric Research* **28**, 57–84.
- Wittchen, H.-U., Lachner, G., Wunderlich, U. & Pfister, H. (1998a). Test–retest reliability of the DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Social Psychiatry and Psychiatric Epidemiology* **35**, 568–578.
- Wittchen, H.-U., Perkonig, A., Lachner, G. & Nelson, C. B. (1998b). Early developmental stages of psychopathology study (EDSP) – objectives and design. *European Addiction Research* **4**, 18–27.
- Wittchen, H.-U., Müller, N. & Storz, S. (1998c). Psychische Störungen: Häufigkeit, psychosoziale Beeinträchtigungen und Zusammenhänge mit körperlichen Erkrankungen. *Das Gesundheitswesen* **60**, 95–100.
- World Health Organization (1990). *Composite International Diagnostic Interview (CIDI), Version 1.0*. World Health Organization: Geneva.