International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 24(2): 116–129 (2015)
Published online 13 October 2014 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/mpr.1455

Measuring symptoms and diagnosing mental disorders in the elderly community: the test–retest reliability of the CIDI65+

HANS-ULRICH WITTCHEN,[1] JENS STREHLE,[1] ANJA GERSCHLER,[1] JANA VOLKERT,[2] MARIA CHRISTINA DEHOUST,[2] SUSANNE SEHNER,[3] KARL WEGSCHEIDER,[3] BERTA AUSÌN,[4] ALESSANDRA CANUTO,[5] MIKE CRAWFORD,[6] CHIARA DA RONCH,[7] LUIGI GRASSI,[7] YAEL HERSHKOVITZ,[8] MANUEL MUNOZ,[4] ALAN QUIRK,[3,6] ORA ROTENSTEIN,[8] ANA BELÉN SANTOS-OLMO,[4] ARIEH SHALEV,[8] KERSTIN WEBER,[5] HOLGER SCHULZ,[2] MARTIN HÄRTER[2] & SYLKE ANDREAS[2,9]

1 Institute of Clinical Psychology and Psychotherapy and Centre of Clinical Epidemiology and Longitudinal Studies (CELOS), Technische Universität Dresden, Dresden, Germany
2 Department of Medical Psychology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
3 Department of Biometry and Epidemiology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
4 School of Psychology, University Complutense of Madrid, Madrid, Spain
5 Division of Liaison Psychiatry and Crisis Intervention, Department of Psychiatry and Mental Health, University Hospitals of Geneva (HUG), Geneva, Switzerland
6 Royal College of Psychiatrists, London, UK
7 Section of Psychiatry, Department of Biomedical and Specialty Surgical Sciences, Ferrara, Italy
8 Department of Psychiatry, Hadassah University Medical Centre, Kiryat Hadassah, Jerusalem, Israel
9 Institute for Psychology, Alpen-Adria Universität Klagenfurt, Klagenfurt, Austria

Key words
reliability, assessment mental disorders, CIDI, anxiety, depression, addiction, elderly

Correspondence
Hans-Ulrich Wittchen, Institute of Clinical Psychology and Psychotherapy and Centre of Clinical Epidemiology and Longitudinal Studies (CELOS), Technische Universität Dresden, Chemnitzer Straße 46, 01187 Dresden, Germany. Telephone (+49) 351-463-36983 Fax (+49) 351-463-36984

Abstract

Prevalence findings for the elderly are artificially low, most likely due to insufficient consideration of age-related cognitive abilities in diagnostic interviews.

Aims: (1) To describe the rationale for the development of an age-adapted Composite International Diagnostic Interview (CIDI65+) for use in a European project (MentDis_ICF65+). (2) To examine its test–retest reliability.

Methods: Based on substantive pilot work the CIDI standard questions were shortened, broken down into shorter subsets and combined with sensitization questions and dimensional measures. Test–retest was determined in $N = 68$ subjects aged 60–79 years via two independent examinations by clinical interviewers using kappa (sensitivity, specificity) for categorical and intraclass correlation (ICC) coefficients for dimensional measures.

Results: Test–retest reliability was good for any mental disorder ($\kappa = 0.63$), major depression ($\kappa = 0.55$), anxiety ($\kappa = 0.62$, range = 0.30–0.78), substance ($\kappa = 0.77$, range = 0.71–0.82), obsessive-compulsive disorder ($\kappa = 1.00$) and most

Email: wittchen@psychologie.tu-dresden.de

Received 22 May 2014;
revised 22 July 2014;
accepted 22 July 2014

core symptoms/syndromes (κ range = 0.48–1.00). Agreement for some disorders (i.e. somatoform/pain) attenuated, partly due to time lapse effects. ICC for age of onset, recency, quantity, frequency and duration questions ranged between $\kappa = 0.60$–0.90. Dimensional agreement measures were not consistently higher.

Conclusion: The age-adapted CIDI65+ is reliable for assessing most mental disorders, distress, impairment and time-related information in the elderly, prompting the need to examine validity.

Introduction

Epidemiological and family genetic studies using standardized diagnostic interviews for mental disorders have consistently demonstrated considerably lower current, 12-month and lifetime prevalence estimates of depressive disorders among the elderly as compared to younger age groups (see Cross-National Collaborative Group, 1992; Knäuper and Wittchen, 1994; Kessler *et al.*, 2010; Jacobi *et al.*, 2013). For example, data of the multisite Epidemiologic Catchment Area (ECA) Study, based on more than 18,000 adults, revealed lifetime prevalence estimates for major depression among persons over the age of 65 of 1.4%, compared to 4.0%, 7.5% and 5.0% for the 45–64, 30–44 and 18–29 year-old groups, respectively (Burke *et al.*, 1991). The most recent nationally representative German epidemiological survey (Jacobi *et al.*, 2013) found a similar gradient not only for depressive disorders but also some other disorders, revealing about two times higher rates for any mental disorder in the youngest (18–34 years old) as compared to the oldest age group (65+); rates for 25–49 and 50–64 year olds were in between. Furthermore, the reported lifetime estimate for every single depressive symptom was found to be lowest among the oldest group (Simon and von Korff, 1992). These findings for depression in particular have however been questioned for several reasons. First, they are counterintuitive, because the elderly have longer time periods of risk for mental disorders, thus one would expect at least higher lifetime rates. Second, clinical experience suggests higher rather than lower prevalence of depressive disorders in the elderly. Further, results of most studies using cross-sectional depression symptom scales suggest increasing rates of depressive symptoms with increasing age (e.g. Berkman *et al.*, 1986; Blazer *et al.*, 1991; Blumenthal, 1975; Gaitz and Scott, 1972; Gurland and Toner, 1983; Klerman, 1988). Consequently there have been efforts to investigate whether finding of lower rates of depression and other mental disorders are fact or artificial.

Focussing on depression, Klerman and collaborators (Klerman and Weissman, 1989; Klerman *et al.*, 1985; Lavori *et al.*, 1987) reviewed a number of potential facts, and several studies have been conducted to investigate some of them empirically (e.g. Hasin and Link, 1988; Andrews *et al.*, 1993; Simon and von Korff, 1992; Warshaw *et al.*, 1991). Alternative explanations include age-related recall bias (i.e. forgetting symptoms experienced earlier in life), sample selection effects of differential mortality and institutionalization. In addition, it has been suggested that the elderly may be less likely than younger subjects to recognize their symptoms as being psychological problems. Blazer (1989) pointed to the potential role of *physical illnesses and conditions* particularly in the assessment of depressive symptoms and disorders in the elderly. He suggested that cases of depression in later adulthood may be missed or labelled incorrectly due to a "masking" of depression by somatic symptoms. Findings regarding this hypothesis are difficult to interpret because assessment strategies in such studies differ according to whether they include or exclude symptoms attributed to physical illnesses. Depression symptom rating scales usually do not differentiate between physical and psychological correlates of depressive symptoms; they count all items to generate an overall score of depressive symptomatology. In contrast, standardized diagnostic interviews, which are designed for assessing mental disorders according to the definitions of DSM-III or DSM-IV [e.g. the Diagnostic Interview Schedule (DIS) (Robins *et al.*, 1981) or the Composite International Diagnostic Interview (CIDI) (Wittchen, 1997; Wittchen and Pfister, 2004)], explicitly exclude symptoms of depression occurring consistently as a result of a physical illness or condition (Blazer *et al.*, 1991; Snowdon, 1990). Thus, one might speculate that the prevalence of depression in later adulthood may be overestimated by the use of symptom depression scales, while the use of diagnostic instruments that incorporate operational diagnostic criteria with their exclusion rules may lead to an underestimate of depression in the elderly (see Newmann, 1989; Scott *et al.*, 2008). Some authors also argued that age-related decrease of prevalence might be a survivor effect, meaning that people with severe mental disorders like depression have a higher mortality rate

(Pinquart *et al.*, 2006; Henderson, 1994). However to test the hypothesis that prevalence declines with age a longitudinal approach in one cohort is needed.

To summarize, there is agreement that the earlier mentioned factors or any combination of them cannot fully explain the low prevalence of depression among the elderly found in population studies (e.g. Klerman and Weissman, 1989; Andrews *et al.*, 1993; Wittchen *et al.*, 1991; Knäuper and Wittchen, 1994). There is also some agreement that there is something special or different about the reporting of symptoms, particularly of those relevant for depression, namely by older respondents and the reporting which seems to be related to the way an assessment instrument takes into account physical illnesses and conditions in the evaluation of symptoms.

Knäuper and Wittchen (1994) examined the possibility that age-related differences observed in epidemiological studies might be due to different ways younger and older respondents process the partly complex symptom questions typically used in standardized diagnostic interviews for mental disorders. They argued that the complex standardized symptom and clinical probe questions require substantial effort and cognitive capacity on the side of the respondent. Subjects have to listen, understand and interpret the content of the stem question; they need to relate the included terms (such as "feeling blue", "lacking energy", etc.) to their individual experience, or more technically, relate to concepts stored in long-term memory. Because the questions require the respondents' judgement about whether the symptom has ever occurred in their lives or certain time frames addressed, respondents have to review their experiences in life, come to a decision and form an appropriate answer. Thus, the evaluation of depressive symptoms is complex and requires a number of cognitive tasks for the respondents (Janca *et al.*, 1992). From a cognitive psychological perspective, the interview tasks involve the parallel storage, retrieval, organization, evaluation and manipulation of complex information, i. e. they require working memory capacity. The theoretical construct "working memory capacity" has been demonstrated to be useful in describing and explaining comprehension and inference processes (see Baddeley, 1986; Daneman and Carpenter, 1980; Just and Carpenter, 1992). Because respondents are asked to review their whole lives, each of the separate cognitive tasks gets more complex with increasing age. Thus, older people need *more* working memory capacity to answer these questions than younger people, but at the same time, ageing is characterized by a decline in working memory capacity (see Salthouse and Babcock, 1991, for a review). Several studies indicated that reductions in working memory capacity explain substantial proportions of age-related variance in verbal processing, e.g. text comprehension (e.g. Burke and Harrold, 1988; Light and Albertson, 1989; Light *et al.*, 1982; Spilich, 1983; Spilich and Voss, 1982; Stine and Wingfield, 1987; Taub and Kline, 1976). We suggest that the required processes of comprehension, memory and judgement are too consuming for older respondents, who need a longer time period to review and may have reduced working memory capacity. Their inability to respond correctly may result in systematic response bias (see Colsher and Wallace, 1991; Rodgers *et al.*, 1988, for related assumptions). Recent research has indicated that with increasing task or question difficulty, respondents are more likely to simplify their task by using more simplistic heuristics, or subjective theories, in constructing a "plausible" answer (e.g. Bless *et al.*, 1992; Bradburn *et al.*, 1987; Nisbett and Wilson, 1977; Reder, 1987). Attributing symptoms to a physical illness or condition might be one frequent strategy used by the elderly to simplify complex recall and judgement processes; it may reflect a particularly "plausible" answer for them. Physically attributed symptoms do not count toward a diagnosis of major depression (Scott *et al.*, 2008). Therefore, the false attribution of symptoms to physical illness may lead to an underestimation of the prevalence of depression.

Such assumed response bias might be less pronounced in more loosely structured clinical diagnostic interviews that lack explicit probe questions. These interviews allow the investigator to use a wider range of flexible and possibly individualized questions in an attempt to adapt the symptom questions to the subject's capabilities. Also, the investigator can use clinical judgement to weight the respondents' answers about whether or not a reported symptom was caused by a physical illness or condition. The use of those clinical interviews, some of which have been developed especially for the assessment of depression in the elderly (e.g. the Comprehensive Assessment and Referral Evaluation (CARE) Depression Scale (Gurland *et al.*, 1977), or the Geriatric Mental State (GMS) Interview (Gurland *et al.*, 1976), frequently results in higher prevalence estimates of major depression (Blazer and Williams, 1980; Kay *et al.*, 1985; but see Henderson *et al.*, 1993, for a review). Fully standardized diagnostic interviews such as the DIS or its follow-up version, the CIDI, do not allow the interviewer to assist the respondent in interpreting the questions. Only the "yes" or "no" answers of the respondent are coded. Thus, the respondent's *subjective judgement* regarding the presence of symptoms and their potential attribution to a physical illness or condition is coded and used for the derivation of a diagnosis. There is also no way for the interviewer to rephrase questions

when the respondent has obvious difficulties in understanding the correct meaning and intent of the question or is clearly misinterpreting the question. Symptoms may be reported as physical when they are psychological, and vice versa (Burvill, 1987; Sandanger, 1993). Fully standardized diagnostic interviews reflect only the respondents' own judgements and perhaps their misunderstanding of questions. Such interviews are, therefore, more vulnerable to attribution bias. Given that most questions require categorical "yes/no" responses, one might argue that if the symptom questions would allow for graded options of a dimensional nature, such response biases might be less pronounced, leading to presumably higher test–retest reliability and validity.

Within a cognitive science perspective, Knäuper and Wittchen (1994) provided some experimental evidence that minor and circumscribed changes in the interview format might help to reduce response biases. They were able to demonstrate: (a) That short and straightforward symptom questions for depression questions with few propositions are reported as frequently in young as well as the older subjects (arguing against an age-related decline of endorsed symptoms). (b) They also found whenever more complex (higher number of propositions) questions or additional probe questions are involved, the frequency of endorsed symptom questions decreases in the elderly. This was especially pronounced when somatic attributions are involved. (c) Further, additional laboratory evidence suggests that "working memory capacity" is a good predictor for this response behaviour, suggesting that complexity of formalized questions might exceed the cognitive capacity of the elderly. The overload might further enhance the probability in the elderly to use causal attributions to somatic illness as a heuristic strategy to solve this problem. Thus, the resulting answer is plausible, but given the intend, incorrect and clearly different from the heuristic used by younger respondents. As a result, the authors suggested to simplify the complex symptom questions used in standardized diagnostic interviews to ultimately increase validity in the elderly.

Aims

This paper is part of a larger cross-national European research project called MentDis_ICF65+ (Andreas *et al.*, 2013) aiming to collect methodological-sound data on the prevalence, incidence, and natural course and prognosis of mental disorders in the elderly. In this paper we describe (1) the developmental process of a modified standardized diagnostic interview of mental disorders for the elderly that incorporates modifications to reduce age-specific response bias in the elderly with the ultimate goal of increasing reliability and validity; (2) we report data on the feasibility of the new instrument and the diagnostic and symptom test–retest reliability when the same person is independently examined twice by two different interviews; (3) we also examine whether a dimensional diagnostic approach to assess depression symptoms results in higher test–retest reliability than a categorical assessment.

Methods

The Composite International Diagnostic Interview (CIDI, DIA-X/CIDI)

The CIDI was chosen as the basis for the instrument development. It is currently the best established assessment platform for mental disorders in the community, and is available in over 16 languages. While there are to our knowledge no reliability studies in older adults, the CIDI approach is reliable at least in younger adults and adolescents in terms of its interrater and test-retest reliability (Reed *et al.*, 1998; Lachner *et al.*, 1998), due to its fully standardized format for use, its computerized administration by clinicians and non-clinicians and the objective computerized diagnostic analysis. Test–retest reliability and procedural validity of the DIA-X/CIDI (Lachner *et al.*, 1998) was tested primarily in a sample of 60 adults (mean age: 22.8 years), with a mean time interval of 39 day between independent examinations. Kappa as a measure for reliability between two raters or two independent examinations indicated good to excellent agreement for any mental disorder ($\kappa = 0.76$), alcohol use ($\kappa = 0.78$) and nicotine dependence ($\kappa = 0.64$), anxiety disorders ($\kappa = 0.81$, range = 0.45–1.00), major depression ($\kappa = 0.68$) and somatoform disorders ($\kappa = 0.62$). Slightly lower, but still acceptable good kappa values were found for eating disorders ($\kappa = 0.56$), mostly due to discrepant duration and frequency reports (Lachner *et al.*, 1998). Originally developed in the context of a World Health Organization (WHO) workgroup in the 1980s (Helzer and Robins, 1984) on the basis of its predecessor the DIS and the Present State Examination (PSE; see Wing, 1996, for a review), the CIDI was subsequently updated, modified and enlarged for use in numerous epidemiological and clinical studies worldwide. In the context of the research programme and for this paper, we used the fully structured algorithm- and computer-based DIA-X/M-CIDI version (Lachner *et al.*, 1998; Reed *et al.*, 1998; Wittchen, 1994; Wittchen and Pfister, 1997) in its computer-assisted version (DIA-X/CIDI) consisting of standardized symptom questions supplemented by self-rating and response lists as part of a mandatory response booklet. The DIA-X/CIDI

allows the fully standardized assessment of symptoms, syndromes and diagnoses for different time frames (four-week, 12-month, and lifetime), along with information about onset, duration, and severity of threshold and subthreshold conditions according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) with its corresponding International Classification of Diseases, 10th revision (ICD-10) F-codes. Additional standard CIDI probe questions allow the description of physical factors and diseases as well as substances that might be causally associated with the symptoms described by the subjects. Diagnoses are derived in a highly objective manner by using exclusively the standardized CIDI diagnostic programme to ensure that the diagnostic criteria are strictly applied on the basis of the symptom information without the interviewer playing any role in making diagnostic statements.

The modification process

The MentDis_ICF65+ CIDI65+ version was developed by the study group under the lead of the first author HUW, who is also a member of the CIDI editorial committee. The process started with an introductory training of all project partners in order to critically review and to identify how the format, the structure and content of the DIA-X/CIDI can be improved for use in the elderly. The general principle in the modification process was to maintain the overall content, structure and rules of the original DIA-X/CIDI as far as possible and to ensure the integrity of all diagnostic variables used for diagnostic computer algorithms (see later for differences).

Diagnostic coverage

Based on pilot work the following syndrome domains and DIA-X/CIDI sections were chosen: nicotine use and dependence (section B), somatoform disorders with an extension covering somatic disease (section C), anxiety disorders (section D), depressive disorders also covering mixed anxiety-depressive syndromes as well as subthreshold forms (section E), bipolar disorders (section F), psychotic symptoms (section G), obsessive-compulsive disorders (section K), adjustment and post-traumatic stress disorders (section N), and cognitive impairment (section M). Alcohol and drug use was assessed in sections I and L using a quantity-frequency screen. Furthermore, the standard DIA-X/CIDI section on help seeking and service utilization (section Q) was retained as well as the section for interviewer observations and clinical overall ratings (section X).

Modifications

Based on the methodological pilot work of Knäuper and Wittchen (1994) the following general and section specific changes were adopted and implemented after pilot-testing:

(a) Complex DIA-X/CIDI symptom questions (as measured by number of propositions) were split into separate subquestions (with no more than six propositions) either explicitly by separate questions, or implictly by interviewer instructions (i.e. Have you ever had a period when you felt sad, blue, depressed? Did this/such periods last two weeks or longer?).
(b) The standard wording of DIA-X/CIDI symptom questions was supplemented particularly in the depression section to include words preferred by the elderly (e.g. depression section: sad and depressed were complemented by down and empty) to ease the respondents symptom recognition.
(c) Whenever appropriate, new sections were opened by either commitment probes, introductory sentences, use of response sheets, dimensional scales or a combination thereof to sensitize the respondent for the respective symptom domain.
(d) The use of the standard DIA-X/CIDI skip-rules were minimized to allow for the assessment of subthreshold conditions (falling short of mandatory DSM-IVTR diagnostic criteria) to allow for subsequent statistical modelling of thresholds.
(e) The socio-demographic DIA-X/CIDI section A was extended and adapted to the core issues of relevance for the elderly.
(f) Dimensional symptom scales were added at the beginning of respective diagnostic sections. We embedded the Hospital Depression and Anxiety Scale (HADS; Zigmond and Snaith, 1983): a 14-item self-rating scale measuring symptoms of depression and anxiety during the previous week prior to the assessment and the Alcohol Use Disorder Identification Test (AUDIT; WHO, 2001): a 10-item self-report questionnaire which is used to identify persons with hazardous and harmful patterns of alcohol consumption. Thus, in advance of the respective DIA-X/CIDI questions, respondents had time to review the presence and the severity of core symptoms. This served the purpose of both sensitization and an option for *post hoc* statistical modelling of alternative threshold definitions of syndromes.
(g) Because of the specific interest in depression, the standard depression screening questions were supplemented by a 29-item screening symptom questionnaire (DSQ-29; Wittchen and Pfister, 2004) presented as part of

Int. J. Methods Psychiatr. Res. 24(2): 116–129 (2015). DOI: 10.1002/mpr

the respondent booklet. Grouped by the nine DSM-IV depression symptom groups, up to five separate symptom questions and synonyms were presented and had to be rated by the respondent for the past four weeks (almost always, about 50% of the time, occasionally, almost never). Any "yes" response of almost 50% of time in any of the nine groups was automatically recorded and counted towards the probability of meeting the respective criteria. Subsequently, the standard DIA-X/CIDI questions were asked. This procedure allowed to compute both the standard CIDI diagnosis as well as a cross-sectional depression score.

(h) To estimate the frequency of suicidal ideation and attempts these questions were separately asked, whenever they were not administered as part of the depression section.

(i) To enrich the elderly specific assessment further self- and clinician-rated instruments (e.g. WHODAS-II, WHO, 2000; WHOQoL-BREF, WHO, 2004), were embedded, that are not part of this paper and described and justified elsewhere (Andreas *et al.*, 2013).

Preparation for test–retest and language versions

The developmental work of the CIDI65+ was performed in parallel in German and English language, the English version served as the benchmark. These two versions were programmed on the basis of the DIA-X/CIDI platform. After extensive pilot-testing and debugging of these two language versions, the respective other language versions (French, Hebrew, Italian) were completed, each first in paper-pencil format and then for computer-assisted personal interviewing (CAPI) use. All translations included back-translation, validation checks and pilot testing. Iterative processes were carried out to improve flow and cross-national translation issues. For the Hebrew CAPI specific solutions (including right–left and characters) were developed.

Training sessions

Centralized and site specific standard CIDI65+ training sessions were conducted overall and for all languages sites. Each course was a standard two-day training session, conducted by one licensed DIA-X/CIDI trainer following the CIDI65+ training guidelines for the DIA-X/CIDI version. Successful CIDI65+ training attendees with two full interviews in real life subjects without major errors were certified by the Dresden centre. A total of 32 trained interviewers were involved in completing the test and the retest interviews, all of which were either psychology or medical postgraduate students.

Design and sample for the test–retest study

We examined in each of the sites convenience samples of inpatients and outpatients with different mental or physical disorders. Each site was asked to sample up to 40 subjects. Subjects were informed about the purpose of the study to assess the quality of the CIDI65+ interview by participating in a test-retest study with ideally a three day time interval between the first and the second interview. Inclusion criteria were age 65+ and written informed consent. Exclusion criteria were severe cognitive impairments [mini-mental state examination (MMSE) cutoff score > 27] making a standardized interview of up to two hours unlikely, no sufficient level of corresponding language and age younger than 64 years.

A total of 302 subjects were approached, 31 did not provide informed consent to the test and 42 denied informed consent for retest interview, 21 did not meet the inclusion criteria. Of the 228 subjects with a test interview only $N = 68$ subjects completed both, the test and retest interview, and were analysed. Reasons for non-completion were predominantly due to not being able to reach the person for the retest interview. Because of this high rate of non-completion, Table 1 compares the 68 subjects analysed with a test and retest interview with those excluded and for which at least minimal information was available. Those non-considered further and those in the analysis did not differ significantly with regard to any socio-demographic variable. Of the $N = 68$ sample analysed, 29 were male, 39 female, mean age was 72.3 years [standard deviation (SD) = 5.7]. The majority of the sample was recruited from outpatient services (55%), about 16% were recruited from inpatient clinics and 29% from other mental health services (e.g. "Rethink" in the UK, which provides services, such as support groups, for people with severe mental health problems and their carers) also including residential homes. The sample size varied across countries, most subjects were interviewed in Dresden (36.8%, $n = 25$) and Geneva (27.9%, $n = 19$), followed by Madrid (14.7%, $n = 10$), Ferrara (10.3%, $n = 7$) as well as Hamburg (7.4%, $n = 5$) and less interviews could have been conducted in London (2.9%, $n = 2$) due to recruitment difficulties (e,g. London: a time-consuming series of research governance procedures to gain access to participants; Jerusalem: technical problem with the CAPI Hebrew language version).

Field procedures

Prior to the first interview respondents were informed that they would be interviewed with the same questions again. It was pointed out that they shall regard both interviews as

Table 1. Socio-demographic characteristics of the MentDis sample (comparison for groups with and without retest)

	Total *N* (%)	No retest *N* (%)	Retest *N* (%)	Comparison		
				χ^2	df	*p*
Gender						
Male	108 (37.2)	79 (35.6)	29 (42.6)			
female	182 (62.8)	143 (64.4)	39 (57.4)	1.11	1	0.292
Age (mean 73.4 years, SD 6.7 years)						
< 65 years	12 (4.1)	8 (3.6)	4 (5.9)			
65–74 years	140 (48.3)	107 (48.2)	33 (48.5)			
75–85 years	115 (39.7)	86 (38.7)	29 (42.6)			
> 85 years	23 (7.9)	21 (9.5)	2 (2.9)	3.64	3	0.303
Marital status						
married	143 (50)	108 (49.5)	35 (51.5)			
widowed	79 (27.6)	57 (26.1)	22 (32.4)			
divorced	40 (14)	33 (15.1)	7 (10.3)			
never been married	20 (7)	17 (7.8)	3 (4.4)			
separated	4 (1.4)	3 (1.4)	1 (1.5)	2.48	4	0.647
Education						
< 9 years of schooling	114 (39.3)	84 (37.8)	30 (44.1)			
9–10 years of schooling	56 (19.3)	47 (21.2)	9 (13.2)			
11–13 years of schooling	62 (21.4)	46 (20.7)	16 (23.5)			
> 13 years of schooling	58 (20)	45 (20.3)	13 (19.1)	2.45	3	0.485
graduated last school “no”	55 (20.1)	41 (19.7)	14 (21.2)			
graduated last school “yes”	219 (79.9)	167 (80.3)	52 (78.8)	0.07	1	0.791
Work status						
retired	241 (84.9)	182 (84.3)	59 (86.8)			
homemaker/housewife	29 (10.2)	24 (11.1)	5 (7.4)			
working/employed	7 (2.5)	4 (1.9)	3 (4.4)			
unemployed	2 (0.7)	2 (0.9)	0 (0)			
Other	5 (1.8)	4 (1.9)	1 (1.5)	2.80	4	0.592
Number of children						
no children	38 (13.1)	31 (14.0)	7 (10.3)			
1–2 children	162 (55.9)	126 (56.8)	36 (52.9)			
3–4 children	46 (15.9)	31 (14.0)	15 (22.1)			
> 4 children	44 (15.2)	34 (15.3)	10 (14.7)	2.83	3	0.418
Self-rated socio-economic status						
below average	48 (16.6)	39 (17.6)	9 (13.2)			
average	82 (28.3)	64 (28.8)	18 (26.5)			
above average	160 (55.2)	119 (53.6)	41 (60.3)	1.11	2	0.573
Self-rated financial situation						
very good	10 (3.5)	101 (47)	2 (2.9)			
Good	112 (39.6)	21 (9.8)	30 (44.1)			
just enough	128 (45.2)	3 (1.4)	27 (39.7)			
Poor	26 (9.2)	8 (3.7)	5 (7.4)			
very poor	7 (2.5)	82 (38.1)	4 (5.9)	5.69	4	0.223

Note: *N*, number; %, percentages; df, degrees of freedom; $p < 0.05$.

independent from each other, as well as that responses to all questions in the second interview should be given regardless of what was reported during the first one. The retest interview was conducted under similar conditions (as far as possible at the same time and in the same premises), but by a different interviewer, who had no information regarding the first CIDI65+ interview to avoid any bias in interviewer behaviour.

Int. J. Methods Psychiatr. Res. 24(2): 116–129 (2015). DOI: 10.1002/mpr

The time interval between the interviews was two to 63 days with a median of seven days and respondents were asked for any critical event, which might have occurred since the first interview prior to the second interview.

All information was collected via the computer-assisted standardized CIDI65+ interview questions and ratings by the respondents or the interviewer. During the data collection a thorough check of the computerized algorithms and adaptation of algorithms, debugging of the CAPI (skip rules, typos, coding instructions) was carried out. Information technology (IT) conventions were agreed and flow of the study processing, including editing procedures was optimized. The programming of the revised data banks for the raw data, behavioural protocol, and automatic plausibility checks were implemented.

Data analysis

Agreement for categorical variables between the test and retest interview was calculated using the kappa statistic

Table 2. Test–retest reliability: Kappa, Yules Y and time effect[1]

		N/T/RT											
CIDI65+ Diagnosis	Total *N*	–/—/+	+/–+/+	%	Sens	Spec	PPV	NPV	"Y"	Kappa	*p*	Beta	*F*-Value
Mental disorders due to SI/GMC	68	55	3	84	0.20	0.95	0.40	0.87	0.36	0.19	0.049	–23.70	0.149
		8	2										
Any substance use disorder	68	56	3	94	0.89	0.95	0.73	0.98	0.85	0.77	< 0.001	42.59	0.096
		1	8										
Nicotine dependence	68	61	3	96	1.00	0.95	0.57	1.00	1.00	0.71	< 0.001	16.65	0.574
		0	4										
Alcohol abuse or dependence	68	61	1	97	0.83	0.98	0.83	0.98	0.89	0.82	< 0.001	20.16	0.575
		1	5										
Any anxiety disorder	68	42	6	84	0.75	0.88	0.71	0.89	0.64	0.62	< 0.001	–3.7	0.823
		5	15										
OCD	68	66	0	99	0.50	1.00	1.00	0.94	1.00	0.66	< 0.001	–22.26	0.660
		1	1										
Any depressive disorder	68	58	4	90	0.50	0.94	0.43	0.95	0.58	0.40	< 0.001	–4.19	0.834
		3	3										
MDD	68	63	1	96	0.50	0.98	0.67	0.97	0.78	0.55	< 0.001	16.17	0.585
		2	2										
Dysthymic disorder (without hierarchy)	68	60	5	90	0.33	0.92	0.17	0.97	0.42	0.17	0.063	–17.21	0.389
		2	1										
PTSD (any A2)	68	18	14	74	0.89	0.56	0.70	0.82	0.52	0.46	< 0.001	–4.19	0.761
		4	32										
Any somatoform disorder	68	63	2	93	0.00	0.97	0.00	0.95	–1.00	–0.04	0.621	–59.21	0.009
		3	0										
Any mental disorder (without nicotine)	68	35	4	82	0.72	0.90	0.84	0.81	0.65	0.63	< 0.001	–9.25	0.562
		8	21										

[1]Time difference between two assessments was regressed on an indicator variable equalling one for observations concordant and zero for observations discordant in disorder measurement (using linear regression).

Note: N, number; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; "Y", Yule's kappa; $p < 0.05$; Beta standardized regression coefficient from linear regression. SI, substance induced; GMC, general medical condition; OCD, obsessive compulsive disorder; MDD, major depressive disorder; PTSD, post-traumatic stress disorder.

(Fleiss and Cohen, 1973; Cohen and Cohen, 1983) and a 12-month time frame for diagnoses. Further we analysed sensitivity, specificity, negative (NPV) and positive predictive value (PPV). Because of the low number of cases in some variables, we additionally report Yules-Y as crude measure of agreement. Kappa (and Yules Y) values of less than 0.40 were considered as poor agreement, values between 0.40 and 0.60 as fair and values above 0.61 and 0.76 as good or respectively excellent agreement. Agreement estimates for continuous variables (e.g. age of onset, duration) were measured using the intraclass correlation (ICC) coefficient. Because of the partly long time lapse between test and retest interview of up to 63 days, that might explain attenuated agreement due to true changes in mental health status, we also estimated the effect of time length for each variable examined. Hereby, the time difference between two assessments was regressed with linear regression on an indicator variable equalling one for observations concordant and zero for observations discordant in measurement of a disorder.

Results

Feasibility and acceptance

The average duration of the complete CIDI65+ administration, including the self-rating scales was 102 minutes (range = 64–143 minutes), the CIDI65 + -specific parts as automatically recorded section by section was 46 minutes (mean; range = 32–68 minutes). The feasibility and acceptance of our modified approach was overall good; three patients asked for a break in between, and two patients refused to answer some questions (both in section N).

Table 3. Anxiety disorders: test–retest reliability: sensitivity, specificity, PPV, NPV, Kappa, Yules Y and time effect[1]

CIDI65+ Anxiety disorder diagnosis	Total *N*	N/T/RT –/—/+	N/T/RT +/–+/+	%	Sens	Spec	PPV	NPV	"Y"	Kappa	*p*	Beta	*F*-Value
Panic attack	68	54	3	84	0.27	0.95	0.50	0.87	0.44	0.27	0.009	–26.06	0.111
		8	3										
Panic disorder	68	55	3	85	0.30	0.95	0.50	0.89	0.47	0.30	0.005	–25.94	0.128
		7	3										
Agoraphobia (DSM-5)	68	63	2	96	0.67	0.97	0.50	0.98	0.78	0.55	< 0.001	–37.50	0.203
		1	2										
GAD	68	63	2	96	0.67	0.97	0.50	0.98	0.78	0.55	< 0.001	–34.33	
		1	2										
Social phobia	68	65	1	97	0.50	0.98	0.50	0.98	0.78	0.48	< 0.001	–32.43	0.367
		1	1										
Any specific phobias	68	50	4	88	0.71	0.93	0.71	0.93	0.70	0.64	< 0.001	–31.58	0.091
		4	10										
Animal type	68	59	2	93	0.57	0.97	0.67	0.95	0.72	0.58	< 0.001	2.51	0.914
		3	4										
Blood–injection–injury type	68	62	1	97	0.80	0.98	0.80	0.98	0.88	0.78	< 0.001	–41.11	0.252
		1	4										
Natural environment	68	64	1	97	0.67	0.98	0.67	0.98	0.84	0.65	< 0.001	–50.27	0.160
		1	2										
Situational type	68	63	1	94	0.25	0.98	0.50	0.95	0.64	0.31	0.004	–68.34	0.007
		3	1										

[1]Time difference between two assessments was regressed on an indicator variable equalling one for observations concordant and zero for observations discordant in disorder measurement (using linear regression).
Note: N, number; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; "Y", Yule's kappa; $p < 0.05$; Beta standardized regression coefficient from linear regression. GAD, generalized anxiety disorder.

Int. J. Methods Psychiatr. Res. 24(2): 116–129 (2015). DOI: 10.1002/mpr

Diagnostic test–retest reliability for main diagnostic groups

In this sample, 35 out of the total of 68 subjects met diagnostic criteria for at least one mental disorder. Based on kappa, respectively, Yules Y in cases with a poor distribution, test–retest reliability for diagnostic groups (Table 2) was overall good for "any mental disorder" ($\kappa = 0.63$). Excellent coefficients were observed for any substance use disorder ($\kappa = 0.77$), any anxiety disorder ($\kappa = 0.62$), and major depression ($\kappa = 0.55$; Y = 0.78). There were no cases with bipolar disorder, thus "any mood disorder" is not indicated. Lower agreement coefficients were found for mental disorders due to medical conditions and substances ($\kappa = 0.19$; Y = 0.36) and dythymia ($\kappa = 0.17$, Y = 42). No significant agreement was found for pain disorders ($\kappa = -0.04$), presumably due to the long time lapse between interviews (significant time effect: $p = 0.009$) found for this disorder only. The specificity and NPV was good to excellent for all disorders (range for sensitivity was 88–98%, range for NPV was 81–100%), sensitivity was fair to good for most disorders, except for mental disorders

Table 4. Test–retest reliability of depression criteria groups in the dimensional and categorical assessment ($N = 68$)

Diagnostic criteria for depression (DSM-IV)	Total *N*	N/T/RT –/—/+	N/T/RT +/–+/+	%	Sens	Spec	PPV	NPV	"Y"	Kappa	*p*
Liste E0 (Dimensional)											
A – Depressed mood	68	51	6	87	0.73	0.89	0.57	0.94	0.65	0.56	< 0.001
		3	8								
B – Loss of interest	68	52	1	90	0.60	0.98	0.90	0.90	0.80	0.66	< 0.001
		6	9								
D – Appetite/weight	68	58	1	91	0.44	0.98	0.80	0.92	0.74	0.53	< 0.001
		5	4								
E – Sleep	68	35	4	79	0.66	0.90	0.83	0.78	0.61	0.57	< 0.001
		10	19								
F – Psychomotor	68	57	3	87	0.25	0.95	0.40	0.90	0.43	0.24	0.021
		6	2								
G – Guilt/worthless	68	57	4	90	0.57	0.93	0.50	0.95	0.63	0.48	< 0.001
		3	4								
H – Concentration/ memory	68	53	3	87	0.50	0.95	0.67	0.90	0.62	0.50	< 0.001
		6	6								
I – Suicidal/ideation	68	59	2	94	0.71	0.97	0.71	0.97	0.79	0.68	< 0.001
		2	5								
CIDI E11-E38 (Categorical)											
A – Depressed mood	68	25	2	88	0.85	0.93	0.95	0.81	0.79	0.76	< 0.001
		6	35								
B – Loss of interest	68	37	6	79	0.68	0.86	0.74	0.82	0.57	0.55	< 0.001
		8	17								
D – Appetite/weight	68	27	7	76	0.74	0.79	0.78	0.75	0.53	0.53	< 0.001
		9	25								
E – Sleep	68	43	6	79	0.58	0.88	0.65	0.84	0.52	0.47	< 0.001
		8	11								
F – Psychomotor	68	38	5	79	0.64	0.88	0.76	0.81	0.57	0.54	< 0.001
		9	16								
G – Guilt/worthless	68	34	3	79	0.65	0.92	0.87	0.76	0.64	0.58	< 0.001
		11	20								

Note: *N*, number; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; "Y" Yule's kappa; $p < 0.05$; Beta, standardized regression coefficient from linear regression.

due to medical conditions and substances (sensitivity 20%), depression (50%) and somatoform disorder (0%).

Diagnostic test–retest reliability anxiety disorders

The test–retest agreement for specific anxiety disorders was good for specific phobias ($\kappa = 0.64$, Y = 0.70; range $\kappa = 0.31$–7.8), good for agoraphobia ($\kappa = 0.55$) and fair to good for panic disorder and generalized anxiety disorder (range Y = 0.47–0.78) (Table 3).

CIDI 65+ stem questions and selected symptoms

In most of the sections, the CIDI65+ presents gateway questions (stem questions) as first items. These stem questions are essential, because negative affirmation leads to skipping of the whole section, and thus, reliability depends heavily on the reliability of these stem questions. Our findings reveal good ($\kappa = 0.51$) to perfect ($\kappa = 1.00$) agreement for the majority of stem questions [tobacco use, somatoform disorders, anxiety, major depression, psychotic disorders, traumatic stress and adjustment disorders (not dealt with in this paper)], while stem question for dysthymia ($\kappa = 0.41$) and (hypo)mania ($\kappa = -0.06$) were lower. Although examinations of individual symptom profile was limited due to small sample size, item analyses for depression revealed mostly moderate to high kappa values of $\kappa = 0.43$ (early awakening) to $\kappa = 1.00$ (suicide attempt).

Continuous measures and age of onset questions

All continuous variables with base rates of four and above showed moderate to high ICC coefficients, ranging between ICC values of 0.771 (number of depressive symptoms) to 0.957 (number of cigarettes). More diversity was found for age of onset questions with high ICCs of above 0.60 for anxiety disorders, depressive disorders and substance use disorders. The notable exception are somatoform disorders (ICC = 0.107). Details available upon request.

Dimensional and categorical assessment of depression symptoms

Table 4 compares the chance corrected kappa agreement between test and retest interview based on dimensional self-report data and the CIDI65+ categorical (yes/no) interview questions, grouped according to the DSM-IV criteria groups for major depression for those domains for which a sufficient number of observations were available. The dimensional assessment preceded the standard categorical CIDI question. The upper part of Table 4 presents the test–retest reliability with the respective kappa values as well as sensitivity, specificity, NPV and PPV. Agreement was examined accordingly for the categorical CIDI65+ questions (lower part of Table 4).

Overall, the dimensional assessment showed a slightly higher agreement (ICC = 0.739) as compared to the categorical assessment (ICC = 0.608). However the individual symptoms agreement was variable. In some items the dimensional assessment appears to be slightly superior (depressed mood, guilt worthlessness and concentration), while in others the dimensional measures appear to be higher (i.e. loss of interest, weight/appetite).

Discussion and conclusion

In this paper we provide preliminary evidence of satisfactory test–retest reliability and the feasibility of the extended and modified CIDI approach. Major limitations are the restricted sample size of the test–retest sample prohibiting the examination of all diagnostic domains, partly low base rates of disorders restricting a more fine graded psychometric exploration, the partly long time lapse between test and retest that significantly impacted the agreement for at least somatoform disorders and the fact that we conducted the study in a convenience sample.

In accordance with substantial previous pilot work (Knäuper and Wittchen, 1994), findings seem to suggest that the major modifications implemented, like shortening of questions by breaking them down in subsets, commitment and sensitivity modules consisting of visual aids and dimensional scales to give respondents more time to reconsider and to remember, implementation of optional synonyms for core symptoms, reduction of skip rules and extensions of dimensional measures are feasible, and resulting in mostly good reliability estimates in the range of previous reliability findings reported for considerably younger samples. Although we are unable to substantiate this via direct comparisons, because such data are not available, we are at least confident, that the format, structure and content of the CIDI65+ fits the needs and capabilities of the elderly better than the standard DIA-X/CIDI. We also conclude that this version can be used in elderly within the same range of limitations as the standard DIA-X/CIDI in younger populations. There are however several caveats: First, we did not test the effect of each of these single measures, thus we do not know which of the modifications made is the most relevant. Second, we were unable to specify to what degree the CIDI65+ provides better reliability than the DIA-X/CIDI, because of the lack of a control group and similarly, unavailability of findings on the test–retest reliability for elderly samples from previous studies. Third, it should be noted that for some

Int. J. Methods Psychiatr. Res. 24(2): 116–129 (2015). DOI: 10.1002/mpr

disorders (e.g. pain and somatoform disorders), reliability is not satisfactory and clearly lower than in previous studies in younger samples.

In previous studies, time intervals between test and retest-interview vary across studies from four to 14 days in the study conducted by Semler *et al.* (1987) and seven to 112 days in the study by Wittchen (1994). Our mean time interval with a median of seven days and a maximum of 63 days in the present study was much longer, and we provided evidence for significant time effects for at least one disorder. Thus, the present reliability coefficients might in fact be higher, because of true changes in the patients' psychopathological state between the two examinations. Although most diagnostic section performed well in this examination, particularly with regard to substance, anxiety and depressive disorders, we confirm previous findings that the assessment of somatoform disorders and disorders due to medical factors and substances are problematic and have no explanation for this problem. This finding might signal the need of further examination of improved strategies particularly in the elderly, because of their frequently high load of somatic morbidity. Further, the reliability coefficients, particularly for depressive disorders and attenuated sensitivity raise concerns. Although some of the discrepancies might partially be explained by the long time lapse, that make it possible that a true change in state might have occurred (e.g. past episode required, incident episode), the agreement coefficients are far from being perfect on both the diagnostic and syndrome level.

Despite frequent claims, particularly within the context of the DSM-5 revision process (Narrow *et al.*, 2009), that dimensional assessment will resolve this problem, our data do not clearly support this idea. Overall, both categorical and dimensional measures performed well, with agreement coefficients for dimensional measures being marginally higher than for categorical measures. However, this difference was limited, and is unlikely to argue for an advantage of dimensional measures in the assessment of depressive disorders. Yet, our study demonstrates at least that an integrated assessment of categorical and dimensional measures in standardized diagnostic interviews is feasible and might have contributed to improved reliability.

The study is an essential first step forward towards improved reliability of standardized diagnostic interviews procedures for mental disorders in older adults. Although we can only speculate whether improved reliability will ultimately also result in improved validity at this point in time, it seems to be fair to state that the CIDI65+ is a reliable diagnostic instrument in terms of its test–retest reliability, facilitating and improving cross-national comparisons of mental disorders in the elderly.

Acknowledgements

This study is funded by a grant from the European Commission (Grant No: 223105) within the Seventh Framework Research Programme of the European Union (EU). The authors gratefully thank all participants and all interviewers of the study.

Declaration of interest statement

The authors have no competing interest to declare.

References

Andreas S., Härter M., Volkert J., Hausberg M., Sehner S., Wegscheider K., *et al.* (2013) The MentDis_ICF65+ study protocol: prevalence, 1-year incidence and symptom severity of mental disorders in the elderly and their relationship to impairment, functioning (ICF) and service utilisation. *BMC Psychiatry*, **13**(1), 62.

Andrews J.A., Lewinsohn P.M., Hops H., Roberts R.E. (1993) Psychometric properties of scales for the measurement of psychosocial variables associated with depression in adolescence. *Psychological Reports*, **73**(3 part 1), 1019–1046.

Baddeley A. (1986) Modularity, mass-action and memory. *Quarterly Journal of Experimental Psychology A*, **38**(4), 527–533.

Berkman L.F., Berkman C.S., Kasl S., Freeman D. H. Jr, Leo L., Ostfeld A.M., Cornoni-Huntley J., Brody J.A. (1986) Depressive symptoms in relation to physical health and functioning in the elderly. *American Journal of Epidemiology*, **124**(3), 372–388.

Blazer D. (1989) Major depression in later life. *Hospital Practice*, **24**(9A), 69.

Blazer D., Burchett B., Service C., George L.K. (1991) The association of age and depression among the elderly – an epidemiologic exploration. *Journal of Gerontology*, **46**(6), M210–M215.

Blazer D., Williams C.D. (1980) Epidemiology of dysphoria and depression in an elderly population. *American Journal of Psychiatry*, **137**(4), 439–444.

Bless H., Mackie D.M., Schwarz N. (1992) Mood effects on attitude judgments: independent effects of mood before and after message elaboration. *Journal of Personality and Social Psychology*, **63**(4), 585–595.

Blumenthal M.D. (1975) Measuring depressive symptomatology in a general population. *Archives of General Psychiatry*, **32**(8), 971–974.

Bradburn N.M., Rips L.J., Shevell S.K. (1987) Answering autobiographical questions: the impact of memory and inference on surveys. *Science*, **236**(4798), 157–161.

Burke K.C., Burke J.D. Jr, Rae D.S., Regier D.A. (1991) Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five US community populations. *Archives of General Psychiatry*, **48**(9), 789–795.

Burke D., Harrold R.M. (1988) Automatic and effortful semantic processes in old age:

experimental and naturalistic approaches. In Ligth L., Burke D. (eds) Language, Memory and Aging, pp. 100–116, New York, Cambridge University Press.

Burvill P.W. (1987) An appraisal of the NIMH Epidemiologic catchment-area programme. *Australian and New Zealand Journal of Psychiatry*, **21**(2), 175–184.

Cohen J., Cohen P. (1983) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 2nd edition, NJ, Erlbaum, Hillsdale.

Colsher P.L., Wallace R.B. (1991) Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology*, **1**(3), 215–230.

Cross-National Collaborative Group. (1992) The changing rate of major depression. Cross-national comparison. *JAMA*, **268**(21), 3098–3105.

Daneman M., Carpenter P.A. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, **19**(4), 450–466.

Fleiss J.L., Cohen J. (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, **33**(3), 613–619.

Gaitz C.M., Scott J. (1972) Age and measurement of mental. *Journal of Health and Social Behavior*, **13**(1), 55–67.

Gurland B., Copeland J., Sharpe L., Kelleher M. (1976) The geriatric mental status interview (GMS). *International Journal of Aging and Human Development*, **7**(4), 303–311.

Gurland B., Kuriansky J., Sharpe L., Simon R., Stiller P., Birkett, P. (1977) The Comprehensive Assessment and Referral Evaluation (CARE) – rationale, development and reliability. *International Journal of Aging and Human Development*, **8**(1), 9–42.

Gurland B., Toner J. (1983) Differentiating dementia from nondementing conditions. *Advances in Neurology*, **38**, 1–17.

Hasin D., Link B. (1988) Age and recognition of depression: implications for a cohort effect in major depression. *Psychological Medicine*, **18**(3), 683–688.

Helzer J.E., Robins L.N. (1984) Lifetime psychotic symptoms assessed with the DIS. *Schizophrenia Bulletin*, **10**(1), 4–7.

Henderson A.S. (1994) Does ageing protect against depression? *Social Psychiatry and Psychiatric Epidemiology*, **29**(3), 107–109.

Henderson A.S., Jorm A.F., MacKinnon A., Christensen H., Scott L.R., Korten A.E., Doyle C. (1993) The prevalence of depressive disorders and the distribution of depressive symptoms in later life: a survey using Draft ICD-10 and DSM-III-R. *Psychological Medicine*, **23**(3), 719–729.

Jacobi F., Mack S., Gerschler A., Scholl L., Hofler M., Siegert J., Burkner A., Preiss S., Spitzer K., Busch M., Hapke U., Gaebel W., Maier W., Wagner M., Zielasek J., Wittchen H.U. (2013) The design and methods of the mental health module in the German Health Interview and Examination Survey for Adults (DEGS1-MH). *International Journal of Methods in Psychiatric Research*, **22**(2), 83–99.

Janca A., Robins L.N., Cottler L.B., Early T.S. (1992) Clinical observation of assessment using the Composite International Diagnostic Interview (CIDI). An analysis of the CIDI field trials – wave II at the St Louis site. *British Journal of Psychiatry*, **160**(6), 815–818.

Just M.A., Carpenter P.A. (1992) A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, **99**(1), 122–149.

Kay D.W., Henderson A.S., Scott R., Wilson J., Rickwood D., Grayson D.A. (1985) Dementia and depression among the elderly living in the Hobart community: the effect of the diagnostic criteria on the prevalence rates. *Psychological Medicine*, **15**(4), 771–788.

Kessler R.C., Birnbaum H., Bromet E., Hwang I., Sampson N., Shahly V. (2010) Age differences in major depression: results from the National Comorbidity Survey Replication (NCS-R). *Psychological Medicine*, **40**(2), 225–237.

Klerman G.L. (1988) The current age of youthful melancholia. Evidence for increase in depression among adolescents and young adults. *British Journal of Psychiatry*, **152**(1), 4–14.

Klerman G.L., Lavori P.W., Rice J., Reich T., Endicott J., Andreasen N.C., Keller M.B., Hirschfield R.M. (1985) Birth-cohort trends in rates of major depressive disorder among relatives of patients with affective disorder. *Archives of General Psychiatry*, **42**(7), 689–693.

Klerman G.L., Weissman M.M. (1989) Increasing rates of depression. Reply. *JAMA*, **262**(7), 900.

Knäuper B., Wittchen H.-U. (1994) Diagnosing major depression in the elderly: evidence for response bias in standardized diagnostic interviews? *Journal of Psychiatric Research*, **28**(2), 147–164.

Lachner G., Wittchen H.U., Perkonigg A., Holly A., Schuster P., Wunderlich U., Turk D., Garczynski E., Pfister H. (1998) Structure, content and reliability of the Munich-Composite International Diagnostic Interview (M-CIDI) substance use sections. *European Addiction Research*, **4**(1–2), 28–41.

Lavori P.W., Klerman G.L., Keller M.B., Reich T., Rice J., Endicott J. (1987) Age-period-cohort analysis of secular trends in onset of major depression: findings in siblings of patients with major affective disorder. *Journal of Psychiatric Research*, **21**(1), 23–35.

Light L.L., Albertson S.A. (1989) Direct and indirect tests of memory for category exemplars in young and older adults. *Psychology and Aging*, **4**(4), 487–492.

Light L.L., Zelinski E.M., Moore M. (1982) Adult age differences in reasoning from new information. *The Journal of Experimental Psychology: Learning, Memory and Cognition*, **8**(5), 435–447.

Narrow W.E., First M.B., Sirovatka P.J., Regier D. A. (eds). (2009) Age and Gender Considerations in Psychiatric Diagnosis. A Research Agenda for DSM-V, Arlington, VA, American Psychiatric Association.

Newmann J.P. (1989). Aging and depression. *Psychology and Aging*, **4**(2), 150–160.

Nisbett R.E., Wilson T.D. (1977) Telling more than we can know – verbal reports on mental processes. *Psychological Review*, **84**(3), 231–259.

Pinquart M., Duberstein P.R., Lyness J.M. (2006). Treatments for later-life depressive conditions: a meta-analytic comparison of pharmacotherapy and psychotherapy. *American Journal of Psychiatry*, **163**(9), 1493–1501.

Reed V., Gander F., Pfister H., Steiger A., Sonntag H., Trenkwalder C., Sonntag A., Hundt W., Wittchen H.-U. (1998) To what degree the Composite International Diagnostic Interview (CIDI) correctly identify DSM-IV disorders? Testing validity issues in a clinical sample. *International Journal of Methods in Psychiatric Research*, **7**(3), 142–155.

Reder L.M. (1987) Strategy selection in question answering. *Cognitive Psychology*, **19**(1), 90–138.

Robins L.N., Helzer J.E., Croughan J., Ratcliff K.S. (1981) National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. *Archives of General Psychiatry*, **38**(4), 381–389.

Rodgers W.L., Herzog A.R., Andrews F.M. (1988) Interviewing older adults – validity of self-

Int. J. Methods Psychiatr. Res. 24(2): 116–129 (2015). DOI: 10.1002/mpr

reports of satisfaction. *Psychology and Aging*, **3**(3), 264–272.

Sandanger I. (1993) The Composite International Diagnostic Interview for psychiatric diagnosis as an expounding instrument in the Norwegian community diagnosis project mental health and physical illness. *International Journal of Methods in Psychiatric Research*, **3**(2), 137–141.

Salthouse T.A., Babcock R.L. (1991) Decomposing adult age-differences in working memory. *Developmental Psychology*, **27**(5), 763–776.

Scott K.M., Bruffaerts R., Simon G.E., Alonso J., Angermeyer M., de Girolamo G., Demyttenaere K., Gasquet I., Haro J.M., Karam E., Kessler R. C., Levinson D., Mora M.E.M., Browne M.A. O., Ormel J., Villa J.P., Uda H., von Korff M. (2008) Obesity and mental disorders in the general population: results from the World Mental Health surveys. *International Journal of Obesity*, **32**(1), 192–200.

Semler G., Wittchen H.U., Joschke K., Zaudig M., von Geiso T., Kaiser S., von Cranach M., Pfister H. (1987) Test–retest reliability of a standardized psychiatric interview (DIS/CIDI). *European Archives of Psychiatry and Neurological Sciences*, **236**(4), 214–222.

Simon G.E., von Korff M. (1992) Re-evaluation of secular trends in depression rates. *American Journal of Epidemiology*, **135**(12), 1411–1422.

Snowdon J. (1990) The prevalence of depression in old. *International Journal of Geriatric Psychiatry*, **5**(3), 141–144.

Spilich G.J. (1983) Life span components of text processing – structural and procedural differences. *Journal of Verbal Learning and Verbal Behavior*, **22**(2), 231–244.

Spilich G.J., Voss J.F. (1982) Contextual effects upon text memory for young, aged-normal, and aged memory-impaired individuals. *Experimental Aging Research*, **8**(3–4), 147–151.

Stine E.L., Wingfield A. (1987) Process and strategy in memory for speech among younger and older adults. *Psychology and Aging*, **2**(3), 272–279.

Taub H.A., Kline G.E. (1976) Modality effects and memory in aged. *Educational Gerontology*, **1**(1), 53–60.

Warshaw M.G., Klerman G.L., Lavori P.W. (1991) Are secular trends in major depression an artifact of recall? *Journal of Psychiatric Research*, **25**(3), 141–151.

Wing J. (1996). SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology*, **31**(2), 50–54.

Wittchen H.-U. (1994) Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. *Journal of Psychiatric Research*, **28**(1), 57–84.

Wittchen H.-U. (1997) DIA-X. Diagnostisches Expertensystem. Interview. Querschnitt, Frankfurt, Swets & Zeitlinger.

Wittchen H.-U., Pfister H. (1997) DIA-X-Interview. Instruktionsmanual zur Durchführung von DIA-X-Interviews, Frankfurt, Swets & Zeitlinger.

Wittchen H.-U., Pfister H. (2004). DIA-X/M-CIDI. DIA-X/CIDI-Composite International Diagnostic Interview. In Strauß B., Schumacher J.. (eds) Klinische Interviews und Ratingskalen, Göttingen, Hogrefe.

Wittchen H.U., Robins L.N., Cottler L.B., Sartorius N., Burke J.D., Regier D. (1991) Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials. *British Journal of Psychiatry*, **159**(5), 645–653, 658.

WHO (World Health Organization). (2000) Disability Assessment Schedule WHODAS-II, Geneva, WHO.

WHO (World Health Organization). (2001) AUDIT/Alcohol Use Disorder Identification Test, Geneva, WHO.

WHO (World Health Organization). (2004) WHOQoL-BREF/WHO Quality of Life HOQoL-BREF, Geneva, WHO.

Zigmond A.S., Snaith R.P. (1983) The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, **67**(6), 361–370.