Pergamon

J. psychiat. Res., Vol. 28, No. 2, pp. 147–164, 1994

0022-3956/94 $7.00 + .00

0022-3956(94)E0007-D

DIAGNOSING MAJOR DEPRESSION IN THE ELDERLY: EVIDENCE FOR RESPONSE BIAS IN STANDARDIZED DIAGNOSTIC INTERVIEWS?

BÄRBEL KNÄUPER* and HANS-ULRICH WITTCHEN

Max Planck Institute for Psychiatry, Clinical Institute, Clinical Psychology, Munich, Germany

(*Received* 6 *August* 1993; *accepted* 14 *January* 1994)

Summary—Recent epidemiological and family genetic studies in different countries using standardized diagnostic interviews for mental disorders have rather consistently demonstrated considerably lower current (e.g. ECA Study: 0.9%) and lifetime (1.4%) prevalence estimates of Major Depression in the elderly (older than 65 years of age) as compared to younger age groups (e.g. 30 44 years: 1 year, 3.9%; lifetime, 7.5%). Some investigators have questioned the validity of these data and suggested alternative interpretations. One possibility is that the complex standardized symptom and clinical probe questions, and the required judgmental processes inherent in diagnostic interviews exceed the cognitive capacity of older adults. This may result in systematic response bias. This paper examines the degree to which the lower prevalence estimates of depression in the elderly are biased due to specific characteristics of the assessment strategy. Analyses of epidemiologic data from the Munich Follow-up Study (MFS), based on the Diagnostic Interview Schedule, demonstrate that (1) older respondents report lifetime depressive symptoms with the same frequency as younger respondents. The additional probe questions designed to identify the degree to which symptoms were caused by factors other than psychological revealed that (2) the elderly more often attribute such symptoms to physical illnesses or conditions. This results in (3) the exclusion of the reported symptoms as a basis for diagnosing depression. A laboratory study demonstrated that "working memory capacity" was a good predictor of this response behavior, indicating that the complexity of the formalized questions exceeds the cognitive capacity of the elderly. Attributing symptoms to a physical illness or condition might be a heuristic strategy to simplify complex recall and judgment processes; the resulting answer is plausible but incorrect. We recommend that the symptom and probe questions of standardized diagnostic interviews be simplified, especially for use with the elderly.

Introduction

SEVERAL recent epidemiological and family genetic studies using standardized diagnostic interviews have rather consistently demonstrated considerably lower current and lifetime prevalence estimates of depressive disorders among the elderly as compared to younger age groups (see Cross-National Collaborative Group, 1992). For example, the data of the multisite NIMH Epidemiologic Catchment Area (ECA) Study (Regier et al., 1984), based on probability samples of over 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

*Author to whom correspondence should be addressed, at: Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106-1248, U.S.A. e-mail:barbel@isr.umich.edv.

5.0% for the 45–64, 30–44 and 18–29 year-old groups, respectively (Weissman et al., 1991). Furthermore, the reported lifetime estimate for every single depressive symptom was found to be lowest among the oldest group (Simon & VonKorff, 1992; Weissman et al., 1991). These findings have been questioned by clinicians, since clinical wisdom suggests higher prevalence of depressive disorders in the elderly. In addition, the results of most studies using depression symptom scales suggest increasing depressive symptomatology with increasing age (e.g. Berkman et al., 1986; Blazer et al., 1991; Blumenthal, 1975; Gaitz & Scott, 1972; Gurland et al., 1983; Klerman, 1988). Alternative interpretations have been suggested, and efforts have been made to assess the possibility that the findings are artifactual.

Possible sources of artifacts

Klerman and collaborators (Klerman & Weissman, 1989; Klerman et al., 1985; Lavori et al., 1987) reviewed a number of potential artifacts, and several studies have been conducted to investigate some of them empirically (e.g. Hasin & Link, 1988; Lewinsohn et al., 1993; Simon & VonKorff, 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. There seems to be agreement, however, that the above 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 & Weissman, 1989; Lewinsohn et al., 1993; Robins & Regier, 1991; Wittchen et al., in press).

Blazer (1989) pointed to another possible confounding factor: the potential role of *physical illnesses and conditions* 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. There is some confusion in the literature about the degree to which physical symptoms (such as loss of appetite and sleep disturbances) are *indicative* of a depressive disorder or, conversely, merely reflect poor physical health and thus should be *excluded* from a diagnosis of depression (see Kessler et al., 1992; Newmann, 1989, for a discussion). Assessment strategies also differ according to whether they include or exclude symptoms attributed to physical illnesses. Depression symptom rating scales do not differentiate between physical and psychological causes of depressive symptoms; they count all these symptoms for a score of depressive symptomatology. In contrast, standardized diagnostic interviews, which are designed for assessing mental disorders according to the definitions of the diagnostic criteria for research DSM-III-R or ICD-10 (e.g. the Diagnostic Interview Schedule (DIS); Robins et al., 1981), explicitly exclude symptoms of depression occurring as a result of a physical illness or condition (Blazer, 1991; Snowdon, 1990). 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). In sum, there is something special or different about the reporting of symptoms of depression by older

respondents related to the degree to which the assessment instrument takes into account physical illnesses and conditions as an exclusion criterion for depressive symptoms. What is lacking in the literature is clear evidence that there are age-related differences in response behavior and that those differences are associated with the attribution of symptoms to physical illness.

Cognitive tasks during probing in the DIS/CIDI

Perhaps, results obtained from standardized diagnostic interviews among the elderly are influenced by age-specific response effects. Low prevalence estimates for the elderly could result from the probe system used to differentiate physically attributed symptoms from possible psychiatric symptoms. This hypothesis can be tested on the basis of cognitive paradigms.

Diagnostic criteria for major depression include prominent and persistent dysphoric mood, loss of interest and the presence of other symptoms (such as changes in appetite and sleep disturbances) nearly every day for at least 2 weeks. These criteria also require clinical judgment because no symptom should be counted towards a diagnosis that is "clearly due to a physical condition" (DSM-III-R, p. 222; American Psychiatric Association, 1987). In diagnostic instruments such as the DIS or the Composite Diagnostic Interview Schedule (CIDI; World Health Organization, 1990), the presence of each criterion is assessed by a series of symptom questions (stem questions). These are followed by different probe questions to assess the degree to which the symptom is entirely explained by physical causes (medication, drugs or alcohol, physical illness, injury or condition). Figure 1 gives a simplified overview of the probe structure.

For every endorsed stem question for a symptom, the respondent is asked whether the symptom was *ever*, and in a second step *always*, caused by a physical illness, injury or condition (or by the use of medication, drugs or alcohol). If a physical origin of the symptom is endorsed, the symptom is coded "4" for illness- or injury-related symptoms. In that case, it does not count towards a diagnosis of depression in the diagnostic algorithm. Only if a physical or other origin of the symptom is negated, or if the respondent or a medical health professional has previously judged the symptoms as being a symptom of a psychiatric illness (or gave an uncertain diagnosis), is the symptom scored as a probable psychiatric symptom. In that case, it is assigned probe code "5". Thus, the probe codes "4" and "5" allow us to distinguish symptoms which, in the respondent's view, were always caused by a physical illness or condition from symptoms of psychogenic origin. It is important to note here that the complete question is not repeated in the probe questions. Only the underlined expressions of the initial stem question are used in these additional probe questions (see Fig. 1).

Thus, the evaluation of depressive symptoms is complex and requires a number of cognitive tasks for the respondents (Robins, 1989). (1) They have to understand and interpret the content of the stem question; they have to relate the included terms (such as "feeling blue," "lacking energy," etc.) to their individual experience or, more technically speaking, to concepts stored in long-term memory. (2) Because the questions require the respondents' judgment about whether the symptom has ever occurred in their lives, respondents have to review their whole lives and search for appropriate episodes in memory

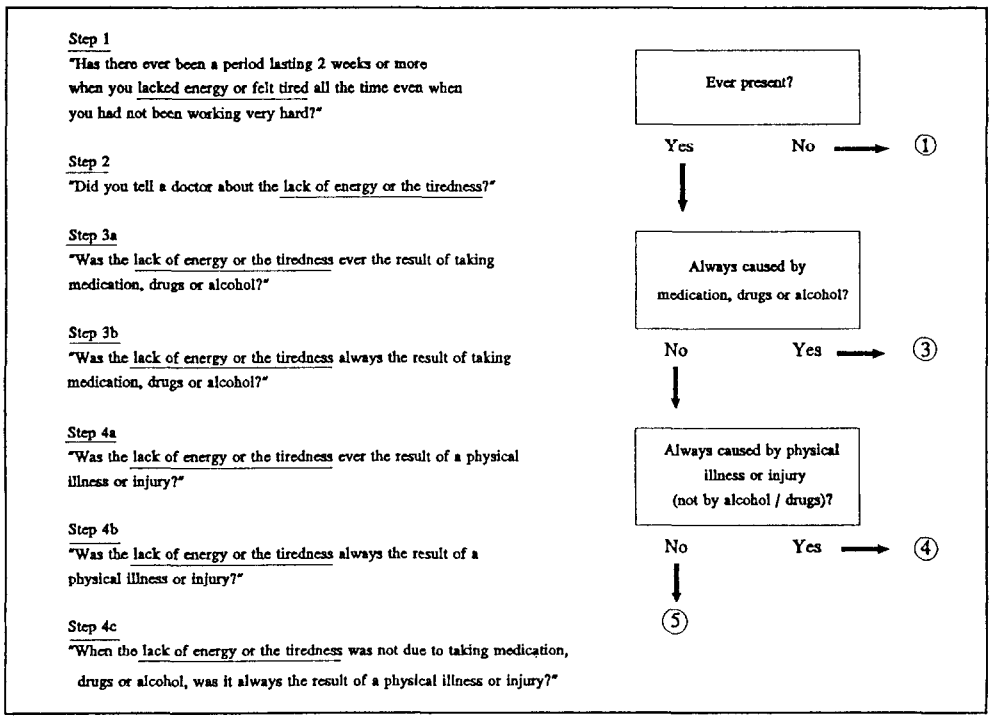

Figure 1. Coding a possible DIS/CIDI depressive symptom (symplified).

Note. 1 = Respondent denies having had the symptom, or does not remember having it; 3 = the respondent has had the symptom and it was always the result of using medication, drugs, or alcohol; 4 = the respondent has had the symptom and it was always the result of a physical illness or injury (or of using medication, drugs, or alcohol); 5 = the respondent has had the symptom and it occurred at least once when it was not explained by either the use of medicines, drugs, or alcohol *or* a physical illness or injury. Thus, this is a possible psychiatric symptom.

("Have you *ever* had ...?"). (3) In doing so, respondents have to consider all details provided in the stem question, that is, all severity and time-related criteria that serve to separate depressive symptoms from less severe and long-lasting distress in everyday life. (4) Finally, respondents have to judge whether the symptoms were ever or always caused by a physical illness or condition or by the use of medication, drugs or alcohol. The severity and time-related criteria of the stem questions are only partially repeated in the subsequent probe questions (see Fig. 1). For example, the time-related criteria "two weeks or more" and "all the time" are not repeated in the probe questions. Respondents are implicitly asked to refer only to symptoms they recalled when responding to the initial symptom question. But to do that, they have to hold all severity and time-related information of the stem question in mind and not refer to evidence other than that already reported. Thus, the complexity of the stem questions might lead to difficulties in accurately answering subsequent probe questions.

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 & Carpenter, 1980; Just & 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 aging is characterized by a decline in working memory capacity (see Salthouse et al., 1991, for a review). Several studies have found that reductions in working memory capacity can explain substantial proportions of age-related variance in verbal processing, e.g. text comprehension (e.g. Burke & Harrold, 1988; Cohen, 1979, 1981; Light & Albertson, 1988; Light et al., 1982; Spilich, 1983; Spilich & Voss, 1983; Stine et al., 1986; Taub & Kline, 1976). We suggest that the required processes of comprehension, memory and judgment are too difficult for older respondents, who have 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 & Wallace, 1989; 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; see Bradburn et al., 1987; Nisbett & Wilson, 1977; Reder, 1987; Schwarz, 1990). Attributing symptoms to a physical illness or condition might be one frequent strategy used by the elderly to simplify complex recall and judgment processes; it may reflect a particularly "plausible" answer for them. Physically attributed symptoms do not count toward a diagnosis of major depression. Therefore, the false attribution of symptoms to physical illness may lead to an underestimation of the prevalence of depression.

This 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 judgment 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 Depression Scale (CARE; Gurland et al., 1977), or the Geriatric Mental State Interview (GMS; Copeland et al., 1976)), frequently results in higher prevalence estimates of Major Depression (Blazer & Williams, 1980; Kay et al., 1985; but see Henderson et al., 1993; see Wittchen et al., in press, 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 judgment* 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 judgments—and perhaps their misunderstanding of questions. Such interviews are, therefore, more vulnerable to attribution bias.

To summarize, we postulate that an age-related reduction in working memory capacity, in combination with the complex cognitive tasks required by standardized diagnostic interview questions and the longer lives the elderly have to review, is associated with a shift away from an exhaustive memory search for the symptoms in questions to a greater tendency toward the computation of "plausible" answers. As a result, symptoms are more likely to be attributed to physical illness. Hence, respondents are not diagnosed as depressed.

Aims

In this paper, we first examine at what step during the diagnostic process of depression age-related differences in response behavior occur. It is assumed that age differences occur predominantly in the frequency with which reported symptoms are attributed to a physical illness or condition. For this purpose, we reanalyzed epidemiological data generated by the 1981 Munich Follow-up Study (MFS; Wittchen & von Zerssen, 1987), based on a representative sample of the western German population. This study also demonstrated lower prevalence estimates of major depression in the oldest as compared to younger age groups (see Cross-National Collaborative Group, 1992). Subsequently, we examine, in an experimental study, the extent to which reductions in working memory capacity can account for the observed age differences in response behavior.

Study 1: Reanalysis of the MFS Data

Methods and population

The analysis is based on data of the follow-up interview of the MFS, a 7 year longitudinal study of a representative German population sample conducted in 1974 and 1981. The basic goals of the MFS were to estimate the prevalence of mental disorders and to provide a comparison of treated vs. untreated cases. A full description of the purposes and methodology of the MFS is given elsewhere (Wittchen et al., 1992; Wittchen & von Zerssen, 1987). A total of 481 subjects, aged 25–64 years at the time of follow-up, was examined in 1981. Diagnostic assessment was based on the German version of the Diagnostic Interview Schedule (DIS; Wittchen & Rupp, 1981). It provides diagnoses according to DSM-III diagnostic criteria (American Psychiatric Association, 1980). For the present analysis of age differences in response behavior, the sample was divided into four 10 year age groups, namely those (1) 25–36 years, (2) 37–46 years, (3) 47–56 and (4) 57–64 years of age.

Results

Frequency of dysphoric mood and reported symptoms (without probing)

Since respondents have to report symptoms for their entire lives, and risk of getting symptoms increases with increasing time, fewer symptoms should be denied in the elderly.

Table 1
Study 1: Frequency (%) of "No"-Answers to Questions of Disphoric Mood and Symptoms in the Eight Symptom Groups of Major Depression by Age (N = 481)

	Age groups			
	(1) 25–36 (n = 105) %	(2) 37–46 (n = 182) %	(3) 47–56 (n = 117) %	(4) 57–64 (n = 77) %
Disphoric mood	80.0	71.5	65.9	71.9
Symptom groups				
Appetite/weight	80.9	81.9	86.3	75.0
Sleep	84.6	78.5	74.7	72.5
Loss of energy/fatigue	65.5	58.1	56.9	63.7
Psychomotor agitation/retardation	90.6	89.8	92.8	83.4
Loss of interest	73.0	60.9	62.8	59.0
Worthlessness/guilt	85.4	83.9	87.0	89.2
Concentration	80.8	81.9	86.0	81.8
Suicidal ideation	83.5	86.7	85.9	88.1

However, this was not the case: no age differences were found in the frequency with which persons of the oldest group denied that they ever in their life experienced dysphoric moods over a period of 2 weeks in comparison to persons of the other age groups (see Table 1; $\chi^2(1, N = 481) = 0.05, p > .50$). There are also no differences in the frequency with which the stem questions of possible depressive symptoms were negated (comparison of the three younger with the oldest group: $\chi^2(1, N = 481) = 0.72, p = .40$). Table 1 presents the frequencies of "no" answers to the stem questions in the eight symptom groups of Major Depression by age. In sum, low estimates of depression in the elderly cannot be attributed to the frequency with which dysphoric mood or other symptoms of depression are reported. Note, however, that the expected decrease in denying of symptoms with increasing age did not appear.

Probing

As described in Fig. 1, stem questions are followed by several probe questions used to determine a possible physical or other cause of the symptom. Symptoms were very seldom attributed to the result of medication, drugs or alcohol, and age differences were non-significant. In the youngest group 1% of the symptoms was attributed to the intake of medication, drugs or alcohol, in comparison to 3%, 0.5% and 2% in the other age groups. However, age differences occurred with regard to the attribution of symptoms to a physical cause. Table 2 presents the percentage of 4 (i.e. physically caused symptoms) and 5 (i.e. not otherwise explained symptoms) probe codes for each symptom for the four age groups. Overall, 4 probe codes increase and 5 probe codes decrease with age. Respondents of the oldest group more frequently report that symptoms were always caused by a physical illness than respondents of the three younger age groups ($\chi^2(1, N = 481) = 14.67, p = .0001$) for all but three symptoms ("poor concentration," "hypersomnia" and "insomnia").

The proportion of the sum of 4 codes to the sum of 5 codes for each symptom is presented

Table 2
Study 1: Frequency (%) of 4 and 5 Probe Codes and Proportion of 4 to 5 for each Symptom by Age (N = 481)

Probe codes Symptoms	Age groups (1) 25–36 (n = 105) %	(2) 37–46 (n = 182) %	(3) 47–56 (n = 117) %	(4) 57–64 (n = 77) %
4 Probe codes				
Psychomotor agitation	0.0	1.2	0.5	5.4
Weight gain	2.2	1.9	2.0	7.8
Poor concentration	1.6	2.2	2.5	2.3
Hypersomnia	1.1	1.2	0.0	0.0
Weight loss	7.0	5.9	9.5	12.4
Decreased libido	0.0	6.2	5.0	4.6
Loss of energy	7.6	7.1	12.0	13.2
Slowed thinking	0.0	0.0	0.0	1.6
Psychomotor retardation	0.5	2.8	3.5	7.8
Loss of appetite	2.7	3.1	3.5	5.4
Insomnia	0.0	1.9	4.5	2.3
5 Probe codes				
Psychomotor agitation	10.8	4.3	1.5	6.2
Weight gain	14.6	8.4	7.0	6.2
Poor concentration	21.1	18.0	13.5	10.1
Hypersomnia	4.9	3.4	2.0	3.1
Weight loss	11.9	9.9	6.5	10.8
Decreased libido	11.9	13.6	12.0	10.1
Loss of energy	22.7	29.5	28.1	17.8
Slowed thinking	8.6	7.4	7.0	7.8
Psychomotor retardation	1.1	3.7	3.0	1.5
Loss of appetite	8.1	9.0	9.5	12.4
Insomnia	15.7	25.7	40.2	37.2
Proportion of 4 to 5				
Psychomotor agitation	0.00	0.28	0.33	0.87
Weight gain	0.15	0.23	0.29	1.26
Poor concentration	0.08	0.12	0.19	0.23
Hypersomnia	0.22	0.35	0.00	0.00
Weight loss	0.59	0.60	1.46	1.15
Decreased libido	0.00	0.46	0.42	0.46
Loss of energy	0.34	0.24	0.43	0.74
Slowed thinking	0.00	0.00	0.00	0.21
Psychomotor retardation	0.46	0.76	1.17	5.20
Loss of appetite	0.33	0.34	0.37	0.44
Insomnia	0.00	0.07	0.11	0.06
Total	0.18	0.26	0.33	0.51

Note. Only symptoms for which a possible physical cause is proved are presented. Percentages are estimated as the sum of 4 probe codes or 5 probe codes, respectively, divided by the total number of responses to the symptom.

in the bottom part of Table 2. For most symptoms, the proportion of 4 to 5 codes is highest in the oldest group. This finding demonstrates that older respondents report general symptoms of depression as frequently as younger respondents. However, during the probing process, they attribute them more often to a physical illness or condition. Physically attributed symptoms are excluded from the further steps of the diagnostic process.

Consequences for criterion B of DSM-III-Major Depression

As a consequence of the higher frequency with which symptoms are attributed to physical causes, a lower proportion of older respondents meets the criterion of four or more "5" coded symptoms. This results in fewer people fulfilling the criterion B of DSM-III Major Depression. Only 6.2% of the oldest group meet this criterion (in comparison to 10.8%, 9.3% and 13.5% in the younger age groups; $\chi^2(1, N = 481) = 4.08$, $p = .04$), and, consequently, fewer people in the oldest group receive a diagnosis of Major Depression.

Study 2: Working Memory Capacity

The findings from the epidemiological study demonstrated that in the elderly (aged 57 or over), fewer symptoms are coded as a depressive (i.e. not otherwise explained) symptom during the probing process. In a subsequent laboratory study, we examined how far reductions in working memory capacity can account for the observed age differences in response behavior.

Method

Subjects

Thirty-one older adults ($M = 63.6$, range 55–75 years) and 32 younger subjects ($M = 31.5$, range 25–40 years), recruited through a newspaper announcement, participated in the study. Half of the subjects in each group were female. No participant reported a serious uncorrected hearing loss. As assessed by the Mehrfachwahl-Wortschatz-Test (MWT-B; Lehrl, 1989), there were no differences between age groups in verbal ability (young, $M = 31.3$, $SD = 0.47$; old, $M = 31.8$, $SD = 0.54$; $t(61) = 0.73$, $p \leqslant .47$). Self-reported current depressive symptomatology as measured by a depressive symptom screening scale (Depressionsskala (D-S'); von Zerssen, 1976) did not differ between age groups (young, $M = 7.3$, $SD = 5.37$; old, $M = 7.9$, $SD = 5.42$, $t(61) = 0.41$, $p > .50$). As assessed by the German short version of the K-scale of the Minnesota Multiphasic Personality Inventory (MMPI-K-Scale; Gehring & Blaser, 1982), there also were no age differences in social desirability scores (both Ms = 12.0, young, $SD = 3.64$; old, $SD = 3.39$). Younger adults had about 2 years more formal education than older adults (young, $M = 13.3$ years, $SD = 0.27$; old, $M = 11.6$ years, $SD = 0.29$; $t(61) = 4.20$, $p \leqslant .0001$). Several aspects of physical health were assessed in the beginning of each session: Subjects were asked about general physical complaints, hearing and vision problems and number and types of medication. Medications were weighted according to the severity of the physical illness they refer to. An index of physical health was created from this information. The index had a range from 0.0 to 9.5 and served as a measure of physical disability, for which we controlled in subsequent analyses. As expected, older persons scored higher on the physical disability index than younger persons (young, $M = 0.66$, $SD = 0.82$; old, $M = 2.98$, $SD = 2.49$, $t(36,20) = -4.95$, $p > .50$, separate variance estimation due to inhomogeneity of variance). More physical health problems were reported with increasing age ($r = .53$, $df = 61$, $p \leqslant .000$).

Working memory span

As a theoretically validated measure of working memory capacity, we used a listening span task, as described by Daneman and Carpenter (1980). This task estimates the ability to hold verbal items in memory while simultaneously performing comprehension, memory search and judgment operations. Several studies have demonstrated the usefulness of the listening span measure as a direct indicator of working memory capacity. In these studies, the working memory capacity of older people has been measured and its correlation with special cognitive abilities (e.g. Salthouse, 1988) and language comprehension abilities (e.g. Baddeley et al., 1985; Hartley, 1986, 1988; Light & Anderson, 1985; Stine & Wingfield, 1987; Wingfield et al., 1988) has been determined. Because a number of these studies confirmed the assumption that this task measures the ability to hold verbal items in mind while performing complex cognitive processes, we used it to estimate the ability of the subjects to perform the complex memory, comprehension and judgment tasks involved in answering the survey questions.

Sixty unrelated true false statements (7–12 words in length; $M = 8.73$, $SD = 1.34$) were constructed. Sentences were selected from the general knowledge domain and were constructed to be of moderate difficulty (e.g. "Cars and bus are examples of vehicles."). The last words of the sentences were different nouns of one to three syllables of length. Items were presented by audiocard to subjects. They responded "true" or "false" to each statement by touching an answer field on a touch screen. Sentences were randomly grouped into sets of two, three, four, five and six sentences. At the end of each set, the subject was instructed to recall the final word of each sentence in that set. Participants were told that the number of sentences would increase after three sets of each length. Subjects began with sets of two sentences and continued until they failed to recall any word in all sets of a level. Ordered recall of the final words was not required. The testing was preceded by two practice sets that contained two items.

The individual working memory span was defined as the highest level at which the participant was able to correctly recall all final words of two of the three sets at a level. Half a point was subtracted if only one out of the three sets of the level was correctly recalled.

Procedure

A computerized version of the depression section of the German version of the CIDI (Wittchen & Semler, 1991) was presented to the participants. The CIDI is a modified, expanded version of the DIS. It uses an almost identical probing procedure to provide diagnoses according to the revised edition of the DSM-III (DSM-III-R; American Psychiatric Association, 1987). To ensure a standardized presentation of the questions without variations in the speed of language or in the feedback behavior of interviewers, the CIDI was presented by audiocard. Several studies have shown that some interviewers tend to adjust their speed and their feedback behavior to the age of respondents while others do not (e.g. Cohen & Faulkner, 1986). Because this would be a source of unsystematic error variance, the speed of language input was standardized. Subjects entered their responses via a touchscreen. The listening span task described above was administered after completion of the interview section.

Results

Frequency of symptoms

Analysis of the age-related response behavior in the laboratory study revealed—even in the comparably small sample—the same results as the secondary analysis of the data of the MFS; no age differences emerged with regard to the frequency of endorsed stem questions of dysphoric mood or loss of interest as well as the other general symptoms of depression (all $\chi^2 < 1$). However, as in Study 1, age differences were found with regard to the attribution of symptoms to physical causes. Older respondents received more 4 probe codes, that is, more often endorsed a physical cause of experienced symptoms, than younger respondents ($\chi^2(1, N = 63) = 3.92, p \leqslant .05$). Consequently older subjects were less likely to qualify for a depressive syndrome (5 codes: $\chi^2(1, N = 63) = 8.38, p \leqslant .004$; see Table 3 for a detailed overview of the findings).

Working memory capacity

The listening span scores were used to examine whether differences in working memory capacity can account for the differences in subjects' responses to the probe questions. The assessment of respondents' working memory capacity with the listening span task revealed a significant negative relationship between age and listening span: With increasing age, the individual listening span of the subjects decreased ($r = -.46, df = 61, p \leqslant .001$). Listening span scores for the age groups were as follows: 25–40 years: $M = 3.64$ words ($SD = 1.29$), 55–64 years: $M = 2.56$ words ($SD = 0.85$), 65–75 years: $M = 2.53$ ($SD = 0.72$). Comparing the young and old group revealed a significant age difference ($t(61) = 3.30, p \leqslant .002$).

Table 3
Study 2: Frequency (%) of 4 and 5 Codes in the Eight Symptom Groups of Major Depression by Age (N = 63)

Symptom groups	Age groups: Young ($n = 32$) %	Old ($n = 31$) %
4 Probe codes		
Appetite/weight	0.8	2.4
Sleep	0.0	2.5
Loss of energy/fatigue	0.0	19.4
Psychomotor agitation/retardation	0.0	3.2
Loss of interest	1.6	3.2
Worthlessness/guilt[a]	—	—
Concentration	0.0	1.1
Suicidal ideation[a]	—	—
5 Probe codes		
Appetite/weight	20.3	11.3
Sleep	35.9	33.9
Loss of energy/fatigue	45.3	32.3
Psychomotor agitation/retardation	7.8	14.5
Loss of interest	31.3	22.6
Worthlessness/guilt	25.0	15.5
Concentration	30.2	33.3
Suicidal ideation	28.9	30.6

[a]For these symptoms a possible physical cause is not proved.

There was a significant negative association between the frequency of "yes" responses to probe questions pertaining to physical cause of the symptom and the working memory span. The shorter the working memory span, the more frequently physical causes were endorsed ($r = -.34$, $df = 61$, $p \leqslant .004$). A two-step hierarchical regression analysis was conducted to test if the age difference in the number of physical attributions can be explained by age differences in working memory. Age was entered into the equation in the first step, followed by the working memory variable in the second step. Table 4 shows the results of these analyses.

The table is organized to show the significance level of the regression coefficients for the variables at each step. The total variance accounted for following each step is also shown. As expected, the regression coefficient of age dropped when working memory span was introduced into the equation. Age is no longer a significant predictor of the response behavior (see Table 4, beta = .18, $p \leqslant .18$). This result confirms our assumption that the age effect can be explained by age differences in working memory. The systematic response behavior seems to reflect the cognitive ability to store and process information in parallel. These results indicate that responses to the probe questions vary systematically with respondents' working memory capacity. Once differences in memory capacity are taken into account, the contribution of age to the variance in "yes" responses to probe questions is low and nonsignificant; the variable "age" loses its predictive power when working memory capacity is taken into account (see Cohen & Cohen, 1975).

Physical health

As described above, subjects reported more physical health problems with increasing age ($r = .53$, $df = 61$, $p \leqslant .001$). Hierarchical regression analyses were used to control for physical health. This was done to insure that the response behavior of attributing symptoms to physical causes is not a function of the actual physical health of the subjects. Priority of entry was again given to the variable "age" while physical health was entered into the equation in the second step. Working memory span was introduced in the third step. The results are presented in Table 5. The regression coefficient of age is not much weakened when physical health is controlled (first step beta of age = .30, $p \leqslant .02$; second step beta = .26, $p \leqslant .08$). However, the regression coefficients of age and physical health substantially drop when working memory span is introduced into the equation in the third step. That is, when age and physical health are controlled for, working memory span remains the only significant predictor of the number of "yes" responses to probe questions of physical causes

Table 4
Study 2: Hierarchical Regression of Age and Working Memory Span on Number of "Yes" Responses to Probe Questions (N = 63)

Predictor	Beta	Significance	R^2 change	R^2
Step I				
Age	.30	$F(1,61) = 6.19$, $p \leqslant .02$	.09	.09
Step II				
Age	.18	$F(1,61) = 1.82$, $p \leqslant .18$	.09	.15
Working memory	−.27	$F(2,60) = 3.99$, $p \leqslant .05$	.06	

Table 5
Study 2: Hierarchical Regression of Age, Physical Health and Working Memory Span on Number of "Yes" Responses to Probe Questions (N = 63)

Predictor	Beta	Significance	R^2 change	R^2
Step I				
Age	.30	$F(1,61) = 6.19, p \leqslant .02$	.09	.09
Step II				
Age	.26	$F(1,61) = 3.16, p \leqslant .08$	.09	.10
Physical health	.09	$F(2,60) = 0.62, p \leqslant .54$	.01	
Step III				
Age	.13	$F(1,61) = 0.75, p \leqslant .39$	.09	.15
Physical health	.09	$F(2,60) = 0.41, p \leqslant .52$	.01	
Working memory	−.27	$F(3,59) = 3.95, p \leqslant .05$	.05	

(beta = .27, $p \leqslant .05$). The results demonstrate that the effect of more frequent reports of a physical cause of symptoms is independent of the current physical health status of the respondents.

In sum, these results confirm the assumption that attributing symptoms to physical illness is a heuristic people use to simplify complex answering processes when they do not have enough working memory capacity to cope with them.

Discussion

Our findings point to the critical role of age-specific response bias in the assessment of depression with standardized diagnostic interviews, such as the DIS or the CIDI. In the elderly, the complex question and probe process results in systematic response bias due to limitations in working memory in the elderly. We show in two independent data sets, one from an epidemiological and the other from a laboratory study, that younger and older subjects report depressive symptoms and major depressive syndromes with similar frequency, provided that the physical condition criterion of DSM-III-R *is not* taken into account. This may help to explain the findings reported by Simon and VonKorff (1992). These authors observed for data of the ECA study that lifetime estimates for every depressive symptom were lowest among the oldest group. That study counted as depressive symptoms only those that fulfilled the exclusionary criterion of a physical condition and thus may have overlooked the critical role of this criterion among the elderly. The lower prevalence estimates for depression in older age groups might, at least in part, be due to the more frequent attribution of symptoms to a physical illness or condition. (These results should, however, be validated by analyzing other existing epidemiological data sets, e.g. the ECA results.)

Furthermore, we have demonstrated that the frequency of attribution of symptoms to physical causes is not restricted to respondents with poor current physical health. This observation is consistent with the assumption that the attribution of reported symptoms to physical causes is a function of response bias due to limitations in working memory processing. This is further supported by data not presented here (see Knäuper, 1994) that revealed systematic age-related differences in response behavior as a function of the *comprehensibility* of the questions: Probe questions that followed stem questions of high

verbal complexity were more often endorsed by older respondents than probe questions that followed stem questions of low verbal complexity. Hence, symptoms assessed by complex stem questions were more often attributed to physical causes than symptoms assessed by stem questions of lower complexity. The response behavior of younger respondents revealed no systematic variation in relationship to question complexity.

The probe system of the DIS and the CIDI requires the respondent to keep all constraints given in the stem question (i.e. time and severity information) in mind, to recall possible symptoms, and to compare the features of both. If the older respondents, due to limitations in working memory, forget that they should review their whole life and should refer only to severe and long-lasting symptoms, or if an exhaustive memory search exceeds their working memory capacity, they may truncate the memory search and may instead refer to a subset of more recent and therefore more accessible symptoms, at the expense of earlier occasions. For these more accessible recent and not severe and long-lasting symptoms, a physical cause is plausible for older subjects because of the common belief and increased probability that aging is characterized by physical distress. Ross (1989), for example, in his research on autobiographical memory, assumes that in recalling events, the more accessible *present status* is used as a benchmark in reconstructing the past and is then combined with *subjective theories* for the respective content domain. In combination, these cognitive processes are likely to result in the construction of a plausible, but incorrect, answer that attributes symptoms to physical causes.

The described response bias effect might also explain why depression scales do not demonstrate a declining depressive symptomatology in the elderly and why some clinical instruments developed for the assessment of depression in the elderly, such as the CARE-System (Gurland et al., 1977) or the GMS (Copeland et al., 1976), suggest higher depression prevalence estimates in the elderly. Obviously, these interviews, which do not use standardized probe questions, allow the investigator to use a wider range of flexible and individualized questions and to weight the respondent's answers by using clinical judgment, thus arriving at lower rates of exclusion of symptoms due to cause by a physical condition.

We can assume that such response bias occurs predominantly in highly standardized diagnostic interviews such as the DIS and the CIDI. Whether this is the case for other standardized interviews has to be examined in future studies. Henderson and collaborators (Henderson et al., in press), for example, reported low rates of depression in a sample of persons over 70 years of age, using a structured instrument that has *no clinical probes* (the "Canberra Interview for the Elderly"; Social Psychiatry Unit, 1992). Thus, a high frequency of discounted symptoms because of physical illness cannot explain the low prevalence estimates in their sample. Also, our identified response bias cannot explain the low depression prevalence estimates found in studies using SADS-L or family history informant data (see Cross-National Collaborative Group, 1992, for a review of results in different countries). Further, the reduced prevalence estimates in the elderly applies to other disorders as well, even when there is no "4" coding option in the respective section of the DIS or CIDI. For example, rates of mania and alcohol abuse are very low. However, it is possible that other disorders (such as alcohol abuse) are *really* less prevalent in the elderly or it may be that other response biases are operating in these sections of the interview. In sum, further research should be conducted to separate "true" from artificial evidence and to identify

possible response biases in retrospective data in addition to that identified in the present study. In general, the use of heuristics or simplifying answering strategies is not restricted to older or cognitively impaired respondents. Research indicates that the use of simplifying answering strategies generally increases with decreasing motivation and increasing time pressure (see Schwarz, 1990, for a review). This emphasizes the general importance of response bias in standardized diagnostic interviews.

Several solutions to the identified problems seem feasible. The first one, of course, is a simplification of the stem and probe questions of standardized diagnostic interviews. A study should be done in which depression rates are compared for DIS- or CIDI-interviews with and without probing. Secondly, memory aids could be used to remind and encourage the older respondent to think seriously about their entire lifetime and to engage in an exhaustive memory search for severe and long-lasting symptoms (cf. Kessler & Wethington, 1991). Moreover, a decomposition of the life review process into smaller time units could improve recall (cf. Bradburn et al., 1987; Kessler et al., 1993). Finally, the probe questions may be used for complete depressive episodes rather than for individual symptoms. This procedure (which has recently been used in the National Comorbidity Study; Kessler et al., 1993) seems to be especially promising. In that case, the assessment of depressive symptoms itself is not confounded with the second step of excluding those episodes that might be caused by a physical illness. Future studies should examine whether age differences in response behavior can be reduced by changing the structure of the probe system.

Acknowledgements—The reported research was supported by grant Wi 709/3-1 from the Deutsche Forschungsgemeinschaft. The authors would like to thank Scott Henderson, Gabi Lachner, Ulrich Mayr, Lee Robins and Norbert Schwarz for helpful comments on a draft of this paper and Tom Trübestein for his assistance.

References

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised). Washington, DC: American Psychiatric Association.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D., Logie, R. Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, **24,** 119–131.

Berkman, L. F., Berkman, C. S., Kasl, S., Freeman, D. H., 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,** 372–388.

Blazer, D. (1989). The epidemiology of depression in late life. *Journal of Geriatric Psychiatry*, **22,** 35–52.

Blazer, D. (1991). Clinical features in depression in old age: A case for minor depression. *Current Opinion in Psychiatry*, **4,** 596–599.

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: Medical Sciences*, **46,** M210–M215.

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

Bless, H., Bohner, G., Hild, T., & Schwarz, N. (1992). Asking difficult questions: task complexity increases the impact of response alternatives. *European Journal of Social Psychology*, **22,** 309–312.

Blumenthal, M. D. (1975). Measuring depressive symptomatology in a general population. *Archives of General Psychiatry*, **137,** 439–444.

Bradburn, N., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, **236,** 157–162.

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

naturalistic approaches. In L. L. Light & D. M. Burke (Eds.), *Language, memory and aging* (pp. 100–116). New York: Cambridge University Press.

Burvill, P. W. (1987). Review: An appraisal of the NIMH epidemiologic catchment area program. *Australian and New Zealand Journal of Pychiatry*, **21,** 175–184.

Cohen, G. (1979). Language comprehension in old age. *Cognitive Psychology*, **11,** 412–429.

Cohen, G. (1981). Inferential reasoning in old age. *Cognition*, **9,** 59–72.

Cohen, J. & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, G. & Faulkner, D. (1986). Does "elderspeak" work? The effects of intonation and stress in comprehension and recall of spoken discourse in old age. *Language and Communication*, **6,** 91–98.

Colsher, P. L. & Wallace, R. B. (1989). Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses. *Journal of Gerontology: Psychological Sciences*, **44,** 45–52.

Copeland, J. R. M., Kelleher, M. J., Kellett, J. M., Gourlay, A. J., Gurland, B. J., Fleiss, J. L., & Sharpe, L. (1976). A semi-structured clinical interview for the assessment of diagnosis and mental state in the elderly: The Geriatric Mental State Schedule. I. Development and reliability. *Psychological Medicine*, **6,** 439–449.

Cross-National Collaborative Group (1992). The changing rate of Major Depression: Cross national comparisons. *Journal of the American Medical Association*, **268,** 3096–3105.

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

Gaitz, C. & Scott, J. (1972). Age and the measurement of mental health. *Journal of Health & Social Behaviour*, **13,** 55–67.

Gehring, A. & Blaser, A. (1982). *Minnesota Multiphasic Personality Inventory: MMPI; deutsche Kurzform für Handauswertung: Handbuch.* Bern: Huber.

Gurland, B., Copeland, J., Kuriansky, J., Kelleher, M., Sharpe, L., & Dean, L. L. (1983). *The mind and mood of aging: mental health problems of the community elderly in New York and London.* New York: Hawthorne Press.

Gurland, B., Kuriansky, J., Sharpe, L., Simon, R., Stiller, P., & Birkett, P. (1977). The comprehensive assessment and referral evaluation (CARE)—Rationale, development and reliability: Part II. A factor analysis. *International Journal of Aging and Human Development*, **8,** 9–42.

Hartley, J. T. (1986). Reader and text variables as determinants of discourse memory in adulthood. *Psychology and Aging*, **1**(2), 150–158.

Hartley, J. T. (1988). Aging and individual differences in memory for written discourse. In L. L. Light & D. M Burke (Eds.), *Language, memory, and aging* (Vol. 3, pp. 36–57). New York: Cambridge University Press.

Hasin, D. & Link, B. (1988). Age and recognition of depression: Implications for a cohort effect in Major Depression. *Psychological Medicine*, **18,** 683–688.

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 late life: A survey using draft ICD-10 and DSM-III-R. *Psychological Medicine*, **23,** 719–729.

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. K., Henderson, A. S., Scott, L. R., Wilson, J., Rickwood, D. J., & Grayson, D. (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,** 771–788.

Kessler, R. C., Foster, C., Webster, P. S., & House, J. S. (1992). The relationship between age and depressive symptoms in two national surveys. *Psychology and Aging*, **7,** 119–126.

Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G., & Nelson, C. B. (1993). Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity, and recurrence. *Journal of Affective Disorders* **29,** 85–96.

Kessler, R. C. & Wethington, E. (1991). The reliability of life event reports in a community survey. *Psychological Medicine*, **21,** 723–738.

Klerman, G. L. (1988). The current age of youthful melancholia. *British Journal of Psychiatry*, **152,** 4–14.

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

Klerman, G. L. & Weissman, M. M. (1989). Increasing rates of depression. *Journal of the American Medical Association*, **261,** 2229–2235.

Knäuper, B. (1994). *Depressions diagnostik im Alter: Verständnis und Verständlichkeit standardisierter diagnostischer Interviewfragen.* Regensburg: Roderer.

Lavori, P. W., Klerman, G. L., Keller, M. B., Reich, T. R., 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,** 23–26.

Lehrl, S. (1989). *Mehrfachwahl-Wortschatz-Test: MWT-B: Manual.* Erlangen: Straube.

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. A. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology*, **102,** 110–120.

Light, L. L. & Albertson, S. A. (1988). Comprehension of pragmatic implications in young and older adults. In L. L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (Vol. 8, pp. 133–153). New York: Cambridge University Press.

Light, L. L. & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology*, **40,** 737–747.

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

Newmann, J. P. (1989). Aging and depression. *Psychology and Aging*, **4**(2), 150–165.

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

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

Regier, D. A., Myers, J. K., Kramer, M., Robins, L. N., Blazer, D. G., Hough, R. L., Eaton, W. W., & Locke, B. Z. (1984). The NIMH epidemiologic catchment area (ECA) program: Historical context, major objectives, and study population characteristics. *Archives of General Psychiatry*, **41,** 934–941.

Robins, L. N. (1989). Diagnostic grammar and assessment: Translating criteria into questions. *Psychological Medicine*, **19,** 57–68.

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,** 381–389.

Robins, L. N. & Regier, D. A. (Eds.) (1991). *Psychiatric disorders in America.* New York: The Free Press.

Rodgers, W. L., Herzog, A. R., & Andrews, F. M. (1988). Interviewing older adults: Validity of self-reports of satisfaction. *Psychology and Aging*, **3,** 264–272.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, **96,** 341–357.

Salthouse, T. A. (1988). The role of processing resources in cognitive aging. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 185–239). New York: Springer.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging*, **6,** 118–127.

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

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (Review of Personality and Social Psychology, Vol. 11, pp. 98–119). Beverly Hills, CA: Sage.

Simon, G. E. & VonKorff, M. (1992). Reevaluation of secular trends in depression rates. *American Journal of Epidemiology*, **135,** 1411–1422.

Social Psychiatry Unit (1992). The Canberra Interview for the Elderly: A new field instrument for the diagnosis of dementia and depression by ICD-10 and DSM-III-R. *Acta Psychiatrica Scandinavica*, **85,** 105–113.

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

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

Spilich, G. J. & Voss, J. I. (1983). Contextual effects upon text memory for young, aged-normal, and aged-impaired individuals. *Experimental Aging Research*, **8**(3), 137–146.

Stine, E. A. L. & Wingfield, A. (1987). Levels upon levels: Predicting age differences in text recall. *Experimental Aging Research*, **13**(4), 179–183.

Stine, E. A. L., Wingfield, A., & Poon, L. W. (1986). How much and how fast: Rapid processing of spoken language in later adulthood. *Psychology and Aging*, **1,** 303–311.

Taub, H. A. & Kline, G. E. (1976). Modality effects and memory in the aged. *Educational Gerontology*, **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,** 141–151.

Weissman, M. M., Bruce, M. L., Leaf, P. J., Florio, L. P., & Holzer, C. (1991). Affective disorders. In L. N.

Robins & D. A. Regier (Eds.), *Psychiatric disorders in America: The epidemiologic catchment area study* (pp. 53–80). New York: The Free Press.

Wingfield, A., Stine, E. A. L., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research*, **14**(2), 103–107.

Wittchen, H.-U., Essau, C. A., Zerssen, D. von, Krieg, J.-Ch., & Zaudig, M. (1992) Lifetime and six-month prevalence of mental disorders in the Munich Follow-Up Study. *European Archives of Psychiatry and Clinical Neuroscience*, **241**, 247–258.

Wittchen, H.-U., Knäuper, B., & Kessler, R. C. (in press). Depression across the life cyle. *British Journal of Psychiatry*.

Wittchen, H.-U. & Rupp, H. U. (1981). *Diagnostic Interview Schedule. German Version* 2. Max Planck Institute for Psychiatry, Munich.

Wittchen, H.-U. & Semler, G. (1991). *Composite International Diagnostic Interview* (*CIDI, Version* 1.0). (*a*) *CIDI-Manual*: *Einführung und Durchführungsbeschreibung*, (*b*) *Interviewheft*. Weinheim: Beltz.

Wittchen, H.-U. & Zerssen, D. von (1987). *Verläufe behandelter und unbehandelter Depressionen und Angststörungen*: *Eine klinisch-psychiatrische und epidemiologische Verlaufsuntersuchung*. Berlin: Springer.

World Health Organization (Eds.) (1990). *Composite International Diagnostic Interview* (*CIDI*): (*a*) *CIDI-Interview* (*Version* 1.0), (*b*) *CIDI-User Manual*, (*c*) *CIDI-Training Manual*, (*d*) *CIDI-Computer Programs*. Geneva: World Health Organization.

Zerssen, D. von (1986). Clinical Self-Rating Scales (CSR-S) of the Munich Psychiatric Information System (PSYCHIS München). In N. Sartorius & T. A. Ban (Eds.), *Assessment of depression* (pp. 270–303). Berlin: Springer.