Sleep Research Online 5(1): 13-43, 2003
http://www.sro.org/2003/Fulda/13/

1096-214X

Cognitive Dysfunction in Sleep-related Breathing Disorders: A Meta-analysis

Stephany Fulda* and Hartmut Schulz**

*Krankenhaus der Barmherzigen Brüder, München, Germany
**Department of Neurology, Helios Klinikum Erfurt, Erfurt, Germany

Sleep related breathing disorders (SRBD) are usually associated with impaired daytime functioning. The magnitude of this impairment might vary for different neuropsychological functions. Our objective was to assess cognitive dysfunction in SRBD patients. Different medical and psychological databases (Evidence Based Medicine, Medline, Embase, PsychInfo, PsychLit, The Eric Database, BiblioSleep) were searched (last search, December 2000). The reference lists of articles were checked and several journals and conference proceedings were hand-searched. All observational studies comparing patients with an established diagnosis of SRBD to non-sleep disordered control groups, to clinical control groups, or to population norms on neuropsychological or psychometric performance measures, including computer-assisted tests and driving simulators. We rated the quality of each study according to criteria of external validity, internal validity, statistical validity, and the level of evidence. Outcome measures were classified according to a taxonomy of neuropsychological functions and statistically analyzed using meta-analytical techniques. Fifty-four studies reporting cognitive functioning of SRBD patients were reviewed. A total of 1,635 patients were compared with 1,737 control subjects. Twenty-eight studies provided adequate statistics and were integrated further. SRBD patients showed moderate to large reductions in mental flexibility, visual delayed-memory retrieval, and driving simulation performance (pooled effect size estimates ranged from 0.61 to 0.72). Small to moderate reductions were found for focused and sustained attention, verbal delayed-memory retrieval, verbal fluency and composite measures of general intellectual functioning (pooled effect size estimates ranged from 0.17 to 0.51). No difference was observed for divided attention, concept formation and reasoning, and verbal or visual immediate-memory performance. Data integration was not undertaken in the areas of attention-span and motor functions due to large between-study heterogeneity, and in the areas of perception, alertness, selective attention, vigilance, constructional performance, learning performance, executive functions and verbal and performance IQ measures due to insufficient data. Our conclusions were that cognitive performance of SRBD patients was impaired, yet there are remarkable differences between various neuropsychological functions and subfunctions. The integrated data show convincingly that disordered breathing during sleep is a risk factor for cognitive functioning during the daytime.

CURRENT CLAIM: Patients with sleep-related breathing disorders experience wide-ranging cognitive dysfunction.

Experimental data corroborate the everyday experience that undisturbed sleep of appropriate duration, intensity and consistency is a prerequisite for adequate cognitive functioning, while sleep disorders are frequently associated with impaired daytime functioning. Major corollaries of disturbed sleep are cognitive dysfunction, mood disorders and social impairment. The kind and degree of impairment differs widely between diagnostic groups, and within groups between patients.

The objective of this meta-analysis is to summarize present knowledge on cognitive dysfunction in patients with sleep related breathing disorders (SRBD), an area in which the majority of studies were published. Cognitive dysfunction in other sleep disorders like insomnia, narcolepsy and restless legs syndrome will be reviewed in a separate meta-analysis.

Patients with SRBD experience cognitive dysfunction that is apparent in most areas of neuropsychological functioning (Hudgel, 1989; Kelly et al., 1990; Day et al., 1999). The available evidence was reviewed in three recent publications (Décary et al., 2000; Engleman et al., 1999, 2000). While Décary et al. (2000) summarized study results on cognitive dysfunction in a narrative review, Engleman et al. (1999, 2000) were the first who provided a quantitative overview of effect sizes and integrated results across studies statistically. They used broad neuropsychological categories such as attention and psychomotor tasks, memory and learning, executive and "frontal" tasks. As Décary et al. showed, construct validity of neuropsychological task performance, especially in the area of attentional functions, is not well understood and has led to very different interpretations even for the same task. The aggregation level for neuropsychological task performance in SRBD patients thus remains to be determined empirically. For this reason, we have tried to combine both approaches. In the present review we summarize evidence on cognitive dysfunction in SRBD patients by grouping individual study outcomes according to the well-established taxonomy of neuropsychological functions by Lezak (1995). If summary statistics were available for individual studies, they were further processed for homogenous groups of functions using meta-analytical techniques. This yields measures of between-study heterogeneity and pooled effect sizes for neuropsychological task performance of SRBD patients.

METHODS

Selection Criteria

For the present review we considered all studies that compared cognitive performance in sleep disordered persons to either (i) cognitive performance in adequate non-sleep

Correspondence: Stephany Fulda, Dipl.-Psych., Krankenhaus der Barmherzigen Brüder, Abt. für Geriatrische Rehabilitation, Romanstr. 93, 80639 München, Germany, Tel: 49-89-1797-3305, Fax: 49-89-1797-3302, Email: StephFulda@compuserve.com.

disordered control groups, (ii) clinical control groups without known neuropsychological impairments, or (iii) population norms. Selection criteria were (a) types of studies (all observational studies), (b) types of participants (sleep disordered persons where the sleep disorder had been established according to at least minimum diagnostic criteria; ASDA, 1990), (c) types of comparisons (to non-sleep disordered control groups, to clinical control groups without known neuropsychological impairment, or to population norms), and (d) types of outcome measures (neuropsychological or psychometric performance measures, including computer-assisted tests and driving simulators).

Search Strategy

The following electronic databases were searched from June to December, 2000 (last search October 12, 2000): Evidence Based Medicine for the period of 1974 to May, 2000; Medline for the period of 1966 to November, 2000; Embase for the period of 1989 to October, 2000; PsychInfo for the period of 1987 to September, 2000; PsychLit for the period of 1987 to June, 2000; The Eric Database for the period of 1982 to June, 2000; and BiblioSleep for the period of 1990 to November, 2000.

The following search terms were used: sleep disorder, sleep apnea, OSAS, CSAS, sleep related breathing disorder, SRBD, upper airway resistance syndrome, UARS, snoring, neuropsychological, cognitive, vigilance, attention, memory, performance, driving simulation. Furthermore, the reference lists of articles were checked and several journals and conference proceedings were hand-searched, especially Sleep Research, Volumes 1 to 25, corresponding to the years 1972 through 1996.

Assessment of Study Quality

All studies were evaluated according to criteria of external, internal and statistical validity. External and internal validity were indexed by two key concepts each (see below). The quality criteria of the present review were based upon criteria for evidence-based medicine (EBM) (Clarke and Oxman, 2000); but also consider the non-interventional nature of those studies where randomized allocation to cases and controls was not possible for obvious reasons. Study quality was assessed with respect to the aims of the present review and was based on information provided in the actual publication. No attempt was made to obtain further information from individual authors.

All studies were evaluated according to (a) external validity related to sampling, (b) external validity related to case definition, (c) internal validity with regard to selection bias, (d) internal validity with regard to performance bias, and (e) statistical validity. Validity was judged to be high, satisfactory, undetermined or unsatisfactory with the exception of statistical validity, which was only judged as high, satisfactory or undetermined. A detailed description of the quality assessment is given in Appendix I.

Levels of Evidence

In evidence-based medicine, results of primary and secondary analysis are classified according to ten levels of evidence (1a to 1c, 2a to 2c, 3a, 3b, 4 and 5; Clarke and Oxman, 2000). These levels describe the best available evidence with regard to a particular research question. So far, evidence-based medicine has not developed standard quality criteria for non-randomized or non-interventional studies that are not concerned with therapeutic interventions, prognosis, diagnosis or economic efficiency. Nevertheless, for the present review levels of evidence for primary studies were closely matched to existing levels of evidence:

1. Individual studies with high external, internal and statistical validity;
2. Individual studies with high internal validity and external validity, which is not outright unsatisfactory;
3. Individual studies with internal and external validity, which is not outright unsatisfactory;
4. Individual studies with unsatisfactory internal or external validity.

We classified each individual study according to the levels of evidence specified above. Whenever feasible, the effect of study quality was examined by excluding retrospectively those studies with poor quality from the analysis, to test for stability of pooled effect sizes.

Classification of Outcome Measures

Neuropsychological outcome measures were grouped according to the taxonomy of neuropsychological functions as proposed by Lezak (1995). In all those cases where at least five independent studies were available that compared the performance of sleep-disordered patients with a control group, outcome measures were aggregated by means of meta-analytical techniques (Voyer et al., 1995).

Integration of Outcome Measures: Effect Sizes and Meta-analysis

In experimental research, the main means of evaluating a scientific hypothesis is the statistical test. The p value resulting from a statistical test is the probability that the effect as estimated from the data may have emerged given the null hypothesis (i.e., no effect) is true. This probability, in turn, is a function of (a) the number of observations, (b) the size of the effect, and (c) the relative efficiency of the statistical test used. It is obvious that components (a) and (c) of the system are specific to the design of a particular study and have only marginal relevance when the hypothesis is to be evaluated on substantial (as opposed to technical) grounds.

The central idea of meta-analysis is that an "average" effect size can be estimated by combining all the unrepresentative, scattered effect sizes obtained in small-scale studies into one combined ("big") effect size that describes the central tendency of the whole distribution of study outcomes. In doing so, the focus of attention is shifted from the idea of significance testing (is the effect greater than zero?) to the idea of estimating the size of an effect (how large is it, exactly?). Since outcomes are often measured by a variety of different questionnaires, computer-assisted tests and miscellaneous observations, each with a different response, metric, study

outcomes have to be standardized to make them statistically comparable. The usual way is to transform the means of the outcome variables into a "z-metric" (a distribution with zero mean and unit standard deviation), and then to compute the number of standard deviations by which two group means differ. The standardized mean difference is the effect size. Effect sizes around 0.2 are considered as small, those around 0.5 as medium, and those of 0.8 or greater as large (Cohen, 1992). An illustrative description of effect sizes states that "medium represents an effect size likely to be visible with the naked eye," a small effect size is to be "noticeably smaller, yet not trivial," and large effect sizes are "the same distance above medium as small is below it" (Cohen, 1992, p. 156). A small effect size is equivalent to the difference in height between 15- and 16-year old girls, a medium effect size is equivalent to the difference in intelligence scores between clerical and semiskilled workers, and a large effect size is equivalent to the difference in intelligence scores between college professors and college freshman (Johnson and Eagly, 2000). When combining effect sizes from different studies, the most common weight is the reciprocal variance so that studies that have larger sample sizes are given more weight. Before computing a weighted mean effect size, the homogeneity of the single study effect sizes must be examined to determine whether the studies can be adequately described by a single effect size. The homogeneity statistic evaluates the hypothesis that the effect sizes are consistent across studies and can thus be meaningfully combined. Given a homogenous set of effect sizes, the result of a meta-analysis is a weighted mean effect size for a population of studies, which can be tested statistically. The typical graphical display of the results (see Figures) shows the effect sizes and confidence intervals from each of the single studies and below the weighted mean effect size from all studies combined. If the confidence interval crosses the vertical axis at zero, an effect size is not significant.

In the present study we aggregated data from single studies within basic neuropsychological functions (e.g., memory) on the level of well-defined sub-functions (e.g., immediate memory) if at least five studies could be found for the given function. Since meta-analysis relies on independent observations, effect sizes from studies comparing two patient groups to one control group, or multiple controls groups with one patient group, were averaged so that only strictly independent observations were entered into each analysis. A technical description of the applied methods is given in Appendix II.

RESULTS

Study Description

The literature search in the electronic databanks yielded a total of 308 documents. All abstracts were read and 167 articles in full-text format were selected for further evaluation. Thirty-four of those were selected for the present review (Findley et al., 1986, 1989, 1995, 1999; Lojander et al., 1999; Sauter et al., 2000; Bédard et al., 1991; Verstraeten et al., 1996, 1997; Schulz et al., 1997; Camus et al., 1999; Stone et al., 1994; Klonoff et al., 1987; Roehrs et al., 1995; Borak et al., 1996; Kotterba et al., 1997; Cassel et al., 1989; Kales et al., 1985; Walsleben et al., 1989; Knight et al., 1987; Risser et al., 2000; Berry et al., 1987, 1990; Muñoz et al., 2000; Juniper et al., 2000; Randerath et al., 2000; George et al., 1996; Naëgelé et al., 1995; Barbé et al., 1998; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997; Ingram et al., 1994; Phillips et al., 1994). Hand searching and the checking of references yielded another 20 documents. Fourteen of these were located in conference proceedings (Zozula et al., 1998a, 1998b; Sloan et al., 1989; Bonanni et al., 1999; Naëgelé et al., 1999; Kuo et al., 2000; Pietrini et al., 1998; Chugh et al., 1998; Lauer et al., 1998; Dani et al., 1996; Dinges et al., 1998; Morisson et al., 1997; Van Son et al., 2000; Verstraeten et al., 2000), two in books (Findley et al., 1991; Weeß, 1996), and four in hand-searched journals (Lee et al., 1999; Kotterba et al., 1998; Rohmfeld et al., 1994; Büttner et al., 2000). Ten studies reported multiple patient (Findley et al., 1986; Lojander et al., 1999; Sauter et al., 2000; Bédard et al., 1991; Rohmfeld et al., 1994) or control groups (Findley et al., 1995; Verstraeten et al., 1996, 1997). Two cases where results from two different studies were reported in one publication were treated as separate studies (Findley et al., 1989, 1991). In another case, where the same patient group was compared with two different control groups (Verstraeten et al., 1996, 1997), the data were treated as one study. The final database thus contained 55 studies. A detailed description of all selected studies is provided in Table 1.

Thirty-three studies compared the performance of SRBD patients and control subjects, sampled from a non-complaining population. Eleven studies compared patients with a control group, sampled in the sleep laboratory. One of them (Findley et al., 1995) included a sample of healthy subjects in addition to subjects who were screened for, but did not fulfil criteria for sleep apnea syndrome (SAS). The clinical control groups included treated patients (Schulz et al., 1997), non-apneic snorers (Verstraeten et al., 1997; Chugh et al., 1998), a mixed group of treated patients and non-apneic snorers (Camus et al., 1999), insomniacs (Verstraeten et al., 1996; Stone et al., 1994) and non-apneic patients referred for evaluation of sleep apnea (Findley et al., 1991, 1995), and a group of patients scheduled for bypass surgery (Klonoff et al., 1987). Ten studies compared performance of SRBD patients to population norms (Findley et al., 1986; Lojander et al., 1999; Sauter et al., 2000; Roehrs et al., 1995; Borak et al., 1996; Kotterba et al., 1997; Cassel et al., 1989; Kales et al., 1985; Walsleben et al., 1989; Verstraeten et al., 2000): one study (Bonanni et al., 1999) compared with an unspecified "database group;" one study (Kotterba et al., 1998) compared with a normal control group and population norms; and one (Stone et al., 1994) compared with a clinical control group and population norms.

There were eight definitions of SRBD used within the studies. The type of sleep-related breathing disorder was defined as obstructive sleep apnea syndrome (OSAS) in 29 studies (Findley et al., 1986, 1989, 1999; Lojander et al., 1999; Sauter et al., 2000; Bédard et al., 1991; Verstraeten et al., 1996,

Table 1
Study Descriptions and Quality Parameters

Study		Patients	Compared To	Study Quality EV S	EV CD	IV S	IV P	SV	Level of Evidence
Findley et al., 1986	a	17 OSAS patients, mean age 53, mean ODI4 43	population norms	0	++				2
	b	9 OSAS patients, mean age 51, mean ODI4 86 with hypoxemia*, 3F 23M (complete sample)	population norms* median SaO2_ 90%, awake PaO2_ 75 mm H						
Lojander et al., 1999	a	10 male OSAS patients, mean age 50 (41-60), median ODI4 31 (10-67)	population norms	-	0				4
	b	12 male OSAS patients, mean age 46 (27-53), mean ODI4 45 (22-72)	population norms						
	c	17 male OSAS patients, mean age 51 (43-65), mean ODI4 26 (11-96)	population norm						
	d	10 male conservatively treated OSAS patients, mean age 49 (40-61), mean ODI4 31 (19-68)	population norms						
Sauter et al., 2000	a	15 OSAS patients, mean age 51, mean RDI 27(<40)	norms	-	++				4
	b	15 OSAS patients, mean age 49, mean RDI 70 (>40)	norms						
Bédard et al., 1991	a	10 male OSAS patients, mean age 53, mean AI 21 (11-29)	10 controls matched for age and sex, mean age 50, mean AI 2	-	++	0	0	+	4
	b	10 male OSAS patients, mean age 51, mean AI 69 (>30)							
Findley et al., 1995	a	62 OSAS patients, 9F 53M, mean age 51, mean AHI 51 (>5)	10 age and sex matched volunteers, 2F 8M, mean age 48	0	+	0	-	0	
	b		12 patients referred for evaluation of SAS, 2F 10M, mean age 50, AHI <5	0	++	+/0	0	0	3
Verstraeten et al., 1996		26 OSAS patients, 3F 23M, mean age 54, mean AHI 48 (>10)	22 insomniac patients, 15F 7M, mean age 47, mean AHI 2	-	++	0	0	0	4
Verstraeten et al., 1996		same patients as Ingram et al., 1999	25 heavy nonapneic snorers, 7F 18M, mean age 50, mean AHI 3						
Findley et al., 1989	a	12 OSAS patients, 3F 9M, mean age 50, mean AI 55	12 age and sex matched controls, 3F 9M, mean age 45	0	+	0	0	+	3
	b	6 OSAS patients, 1F 5M, mean age 46, mean AI 83 (ODI4 >50)	7 age and sex matched controls, 1F 6 M, mean age 44	0	+	0	0	+	3
Schulz et al., 1997		16 male OSAS patients, mean age 47, apnea severity unknown	17 male CPAP-treated OSAS patients, mean age 46	-	++	0	0	0	4
Camus et al., 1999		12 OSAS patients, 2F 10M, mean age 52 (40-66), mean REI 71 (28-174)	10 persons with REI <11 (snorers, CPAP treated patients),1F 9M mean age 52 (40-69), mean REI 3 (0-10)	-	++	0	+	++	4
Stone et al., 1994		18 elderly insomniac patients with OSAS, 3F 15M, mean RDI 24 (>10); patients and controls taken from the same sample (mean age 65 (> 55))	16 elderly insomniac patients without OSAS, 14F 2M, RDI <5, comparable on age, IQ and education; age, IQ matched norms	0	++	++	+	+	2
Klonoff et al., 1987		11 male OSAS patients, mean age 49 (34-66), mean AI 49 (>15) before UPP	11 male age-matched patients, mean age 49, before coronary bypass surgery (CBS)	-	+	0	0	0	4
Roehrs et al., 1995		25 male OSAS patients, mean age 49, mean REI 66	population norms (cut-off 2 SD)	0	++				2
Borak et al., 1996		20 male OSAS patients, mean age 46, mean AHI 67	population norms	-	++				4
Kotterba et al., 1997		40 OSAS patients, 1F 39M, mean age 52 (34-74),mean AHI 53 (10-105)	age-matched population norms (cut-off percentile rank <25)	-	++				4
Cassel et al., 1989		22 SAS patients, 2F 20M, mean age 51, mean AI 36 (11-92)	age-corrected population norms	-	++				4
Kales et al., 1985		50 OSAS patients, 6F 44M, mean age 48 (23-68), "apnea severe enough to warrant a recommendation for trachestomy"	population norms	0	++				2
Walsleben et al., 1989		7 of 14 OSAS patients, part of a larger sample with 3F 11M, mean age 52 (37-61), apnea severity unknown but recommended for CPAP	population norms	-	++				4
Knight et al., 1987		10 elderly OSAS patients, 4F 6M, mean age 78 (>64), mean AI 17 (7-38)	17 elderly controls, 12F 5M, mean age 75 (>64)	0	++	+	++	+	2

Table 1 (cont.)

Study	Patients	Compared To	Study Quality					Level of Evidence
			EV		IV		SV	
			S	CD	S	P		
Risser et al., 2000	15 OSAS patients, 2F 13M, mean age 42 (29-51), mean AI 47 (23-98)	15 controls, 6F 9M, mean age 38 (30-50)	-	+	0	0	+	4
Findley et al., 1999	31 OSAS patients, 4F 27M, mean age 45, mean AHI 46 (10-99)	14 volunteers, 3F 11M, mean age 43	-	+	0	-	+	4
Berry et al., 1990	8 male geriatric subjects with SAS, mean age 69, mean AHI 28 (>10)	12 male controls, mean age 68, mean AHI 3	-	++	0	0	0	4
Muñoz et al., 2000	80 SAS patients, 2F 78M, mean age 49, mean AHI 60 (21-123)	80 controls, 2F 78M, mean age 46	0	+	0	0	0	3
Juniper et al., 2000	12 male OSAS patients, median age 48, median ODI4 41 (>10)	12 male controls, median age 49, median ODI4 1	-	0	0	0	+	4
Randerath et al., 2000	28 OSAS patients, 1F 27M, mean age 52, mean AHI 38	52 controls, 9F 42M,, mean age 24	-	+	0	0	+	4
George et al., 1996	21 male OSA patients, mean age 49, mean AHI 73 (>15)	21 male age-matched controls, mean age 46, mean AHI 3 (<15)	0	++	0	+	0	3
Naëgelé et al., 1995	17 male SAS patients, mean age 49 (27-76), mean RDI 41 (14-75)	17 matched controls (age, verbal IQ, school education level), mean age 49	0	+	0	0	0	3
Barbé et al,. 1998	60 OSAS patients, 1F 59M, mean age 47, mean AHI 58 (21-101)	60 controls individually age and sex matched	0	+	0/-	0	0	3
Redline et al., 1997	32 SDB patients, 17F 15M, mean age 51 (40-65), mean RDI 17 (9-27)	20 volunteers, 12F 8M, mean age 49 (40-65), mean RDI 2 (0-5); no difference in age, education and estimated IQ (means adjusted for age, race and estimated IQ)	0	++	0	0	0	3
Greenberg et al., 1987	14 SAS patients, 1F 13M, mean age 44 (<55), mean AI 48	14 healthy volunteers, 3F 11M, mean age 44 (<55), no significant difference on age, premorbid IQ and years of education	-	+	0	-	0	4
Kim et al., 1997	199 SAHS patients, AHI range 5-97 (mean AHI 19); age and gender only known for the complete sample: mean age 45 (30-60), 338F 503M	642 controls, AHI <5; within prospective snorer enriched sample	++	++	++	++	++/0	1
Ingram et al., 1994	16 elderly subjects with AI >5, 6F 10M, mean age 63 (>53), mean AI 19	43 elderly subjects with AI <5, 27F 16M, mean age 62 (>53), mean RDI 1	0	++	+	+	++/0	2
Phillips et al., 1994	13 SBD patients, 5F 8M, mean age 73 (>50), mean AHI 11 (>5)	53 controls, 27F 26M, mean age 67 (>50), mean AHI 1 (<5)	+	++	0	0	+	3
Berry et al., 1987	8 persons with AHI >/=5, mean AHI 15; complete sample mean age 69 (>60), 11F 18M	21 persons with AHI <5, mean AHI 1	+	++	+	++	0	2
Zozula et al., 1998a	8 OSAS patients, 1F 7M, mean age 37, mean AHI 73 (>25)	8 controls, 1F 7M, mean age 40, mean AHI 6	-	++	0	+	0	4
Sloan et al., 1989	32 OSAS patients, 2F 30M, mean age 45, mean apnea severity not reported	19 male controls, mean age 43	-	0	0	0	0	4
Bonanni et al., 1999	18 male OSAS patients, aged 40 to 63, apnea severity not reported	databases	-	++				4
Naëgelé et al., 1999	20 mildly impaired OSAS patients, mean RDI 24; part of a larger sample of 41 patients with mean age 51, 8F 33M	30 controls, mean age 49, 8F 22M, matched for age and IQ	-	+	0	0	0	4
Kuo et al., 2000	53 untreated subjects with RDI >19, mean age 60, mean RDI 33 (20-50); in the complete sample 44F 60M, neuropsychological investigation 1 to 2,5 years after PSG	51 subjects with RDI <5, mean age 58, mean RDI 3	++	++	+	0	0	3
Pietrini et al., 1998	8 OSAS patients, 1F 7M, mean age 50, mean AHI 42	8 matched controls, 1F 7M, mean age 50, mean AHI 1	-	++	0	0	0	4
Chugh et al., 1998	22 OSAS patients, 11F 11M, mean age 43, mean RDI 47 (>10)	10 snorers, 6F 4M, mean age 45, mean RDI 4 (<10)	-	++	0	0	0	4
Zozula et al., 1998b	16 OSAS patients, 5F 11M, mean age 41, AHI >25	age and gender matched controls	-	++	0	0	0	4
Lauer et al., 1998	50 OSAS patients, age range 22 to 68, RDI >10	22 controls matched for age, gender, and years of education	-	+	0	0	0	4
Dani et al., 1996	5 SAS patients, 1F 4M, mean age 53, mean AHI 50	5 controls, 1F 4M, mean age 54, mean AHI 1	-	++	0	0	+	4

Study		Patients	Compared To	Study Quality					Level of Evidence
				EV		IV		SV	
				S	CD	S	P		
Dinges et al., 1998		78 subjects with RDI >5; for the complete sample mean age 49, 6F 192M	120 subjects with RDI < 5	++	++	++	+	0	2
Morrison et al., 1997		23 OSAS patients, 2F 21M, mean age 44, REI >20	20 controls, 5F 15M, mean age 45	-	++	0	0	+	4
Van Son et al., 2000		12 OSA patients, no age, gender or apnea severity reported	10 controls no age or gender reported	-	++	0	0	0	4
Verstraeten et al., 2000		17 OSA patients, mean age 49, mean AHI 41, gender unknown	population norms	-	++				4
Findley et al., 1991	a	50 elderly subjects with AI >5, 7F 43M, mean age 61 (>55), mean AHI 46	21 elderly subjects evaluated for SAS with AHI<5, 3F 18M, mean age 59 (>55), mean AHI 3	0	++	+	+	0	3
	b	36 elderly subjects with AHI>5, 3F 33M, mean age 59, mean AHI 50	15 elderly subjects evaluated for SAS with AHI<5, 2F 13M, mean age 57, mean AHI 2	-	++	0	0	+	4
Weeß, 1996		15 male OSAS patients, mean age 51, mean RDI 43	15 male controls, mean age 49, mean RDI 2	0	++	0	0	0	3
Lee et al., 1999		17 subjects with SAS, 7F 9M, mean age 49, mean RDI 39 (12-85)	16 controls, 9F 7M, mean age 45, mean RDI 2 (<7)	0	++	+	0	0	3
Kotterba et al., 1998		31 male OSAS patients, mean age 50 (34-72), mean AHI 37	10 male volunteers, mean age 48 (36-66); age-corrected population norms	-	+	0	0	0/-	4
Rohmfeld et al., 1994	a	10 male OSAS patients, mean age 51, mean RDI 12 (6-29)	12 male controls, mean age 45, RDI <5	-	++	0	0	+/0	4
	b	8 male OSAS patients, mean age 54, mean RDI 48 (>30)							
Büttner et al., 2000		90 OSAS patients, 16F 74M, mean age 59, mean AHI 34	100 controls, 10F 90M, mean age 34	-	+	-	0	0	4

EV-S: external validity related to sampling; EV-CD: external validity related to case definition; IV-S: internal validity related to selection bias; IV-P: internal validity related to performance bias; SV: statistical validity. A detailed description of validity parameters is given in Appendix 1.

1997; Schulz et al., 1997; Camus et al., 1999; Klonoff et al., 1987; Roehrs et al., 1995; Kotterba et al., 1997, 1998; Kales et al., 1985; Knight et al., 1987; Risser et al., 2000; Berry et al., 1990; Muñoz et al., 2000; Juniper et al., 2000; Zozula et al., 1998a, 1998b; Pietrini et al., 1998; Lauer et al., 1998; Morisson et al., 1997; Weeß, 1996; Rohmfeld et al., 1994; Büttner et al., 2000), obstructive sleep apnea (OSA) in eleven studies (Findley et al., 1991, 1995; Borak et al., 1996; Walsleben et al., 1989; George et al., 1996; Sloan et al., 1989; Bonanni et al., 1999; Chugh et al., 1998; Van Son et al., 2000; Verstraeten et al., 2000), sleep apnea syndrome (SAS) in four studies (Naëgelé et al., 1995; Barbé et al., 1998; Dani et al., 1996; Lee et al., 1999) and occasionally sleep apnea (SA) (Cassel et al., 1989), obstructive sleep apnea/hypopnea syndrome (Naëgelé et al., 1995), sleep disordered breathing (Redline et al., 1997), sleep apnea DOES syndrome (Greenberg et al., 1987), or insomnia with obstructive sleep apnea (Stone et al., 1994). Six studies, in which the groups were identified outside the sleep laboratory, distinguished between cases and controls on the basis of the apnea/hypopnea index (AHI) (Kim et al., 1997; Ingram et al., 1994; Phillips et al., 1994; Berry et al., 1987) or the respiratory disturbance index (RDI) (Kuo et al., 2000; Dinges et al., 1998). Apnea severity indices that were reported included the apnea hypopnea index (AHI, 22 studies), the apnea index (AI, 9 studies), the respiratory disturbance index (RDI, 12 studies), the respiratory event index (REI, 2 studies), or the oxygen desaturation index (ODI, 3 studies). Six studies did not report an apnea severity index. Average apnea severity measures ranged for AHI from 11 (Phillips et al., 1994) to 73 (George et al., 1996; Zozula et al., 1998a), for AI from 17 (Knight et al., 1987) to 83 (Findley et al., 1989), for RDI from 12 (Rohmfeld et al., 1994) to 30 (Sauter et al., 2000), for ODI from 26 (Lojander et al., 1999) to 86 (Findley et al., 1986), and for the respiratory event index (REI) from 66 (Roehrs et al., 1995) to 71 (Camus et al., 1999).

The minimal diagnostic requirement for the diagnosis of sleep-related breathing disorders in the present review was nocturnal oximetry, which was considered to have been performed if an apnea severity index was reported. The majority of studies also performed a full night polysomnography to establish the diagnosis in the patient group, with four exceptions: one study (Lojander et al., 1999) used oximetry in combination with the static-charge-sensitive-bed; one study (Juniper et al., 2000) used oximetry and snoring; and two studies did not specify diagnostic procedures but provided measures of apnea severity (Sloan et al., 1989; Dani et al., 1996). For these four latter studies, external validity related to case definition was considered undetermined. Although not all subjects with sleep-disordered breathing were patients, for the sake of simplicity, we will refer to them as SRBD patients in the following.

The average age varied between 37 and 78 years for the SRBD patients and between 34 and 75 years for the control subjects, with a peak between 40 and 50 years for both groups. Forty-eight studies reported the gender of patients and 39 of them did so for the control group. Thirty studies included females, with a total of 132 females in patient groups and 201 in control groups. In comparison, a total of 995 males were included in patient groups and 669 in control groups. Summarized across all studies, there were 1,635 SRBD patients and 1,737 control subjects.

Study Quality

The results of the evaluation process are summarized in Table 2. External validity related to case definition was high for 35 studies, satisfactory for 16 studies, and undetermined for four studies. External validity related to sampling was high in only three studies, satisfactory in another two, undetermined in 16 studies, and unsatisfactory in 33 studies. Internal validity and statistical validity were only evaluated for those 44 studies, which compared performance of SRBD patients and controls. Internal validity with regard to selection bias was high in four studies, satisfactory in six studies, undetermined in 32 studies, and unsatisfactory in one study. Likewise, internal validity controlling for performance bias was high in three studies, satisfactory in eight studies, undetermined in 30 studies, and unsatisfactory in two studies. Statistical validity was high in three studies, satisfactory in 13 studies, and undetermined in 27 studies. Regarding the level of evidence for single studies only one study was judged as Level 1 evidence, while seven were Level 2, 14 Level 3, and 32 Level 4 evidence.

Neuropsychological Functions

Perception

Four studies investigated basic perceptual abilities in SRBD patients Bédard et al., 1991; Knight et al., 1987; Dani et al., 1996; Lee et al., 1999). Patients did not differ from controls in skin writing perception (graphesthesia) (Knight et al., 1987), the Hooper visual organization test (Bédard et al., 1991), and a visual matching test (Bédard et al., 1991). In a sensory motor task, patients showed a higher number of correct responses than controls (Lee et al., 1999). Finally, Dani et al. (1996) reported reduced facial recognition in a small group of five SRBD patients when compared to controls. No data integration was undertaken since the number of studies was small and the tasks employed diverse.

Attention

Thirty-eight studies have assessed attentional performance (Findley et al., 1986, 1991; Sauter et al., 2000; Bédard et al., 1991; Verstraeten et al., 1996, 1997, 2000; Schulz et al., 1997; Camus et al., 1999; Stone et al., 1994; Roehrs et al., 1995; Borak et al., 1996; Kotterba et al., 1997, 1998; Cassel et al., 1989; Walsleben et al., 1989; Knight et al., 1987; Muñoz et al., 2000; Randerath et al., 2000; Naëgelé et al., 1995; Barbé et al., 1998; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997; Phillips et al., 1994; Zozula et al., 1998a, 1998b; Sloan et al., 1989; Bonanni et al., 1999; Kuo et al., 2000; Pietrini et al., 1998; Chugh et al., 1998; Lauer et al., 1998; Dani et al., 1996; Dinges et al., 1998; Morisson et al., 1997; Weeß, 1996; Lee et al., 1999; Rohmfeld et al., 1994). Since most studies reported multiple outcome measures, data will be reviewed for different areas of attention separately.

Measures of alertness were employed in six studies (Bonanni et al., 1999; Verstraeten et al., 2000; Weeß, 1996; Lee et al., 1999; Kotterba et al., 1998; Rohmfeld et al., 1994). SRBD patients and controls did not differ in the Critical Flicker Fusion test (CFF) in two studies (Weeß, 1996; Rohmfeld et al., 1994) and in a short two-minute choice reaction time task (Lee et al., 1999). Simple reaction time, on the other hand, was prolonged in patients when compared to controls and norms (Kotterba et al., 1998) as well as to an unspecified database (Bonanni et al., 1999). Verstraeten et al. (2000) found that while some patients showed impaired performance in a phasic alertness task, performance in a tonic alertness task was unimpaired in patients when compared to norms. Only three studies reported means and standard deviations, so that no data integration was undertaken.

Attention span was assessed in ten studies in the auditory (Borak et al., 1996; Knight et al., 1987; Naëgelé et al., 1995;

Table 2
Study Quality

Source	Level of Evidence: No. Level	Studies	Study Quality	External Validity: Sampling	External Validity: Case Definition	Internal Validity: Selection Bias	Internal Validity: Performance Bais	Statistical Validity
Article	1	1	High	1	24	2	3	3
	2	6	Satisfactory	2	13	4	4	11
	3	11	Undetermined	15	2	23	21	16
	4	21	Unsatisfactory	21	-	1	2	-
Abstract	1	-	High	2	10	1	-	-
	2	1	Satisfactory	-	2	1	2	2
	3	1	Undetermined	-	2	10	10	10
	4	12	Unsatisfactory	12	-	-	-	-
Book	1	-	High	-	2	-	-	-
	2	-	Satisfactory	-	-	1	2	-
	3	2	Undetermined	1	-	1	-	2
	4	-	Unsatisfactory	1	-	-	-	-

Redline et al., 1997; Greenberg et al., 1987; Pietrini et al., 1998; Lauer et al., 1998; Dani et al., 1996; Verstraeten et al., 2000; Lee et al., 1999) and visual domain (Naëgelé et al., 1995; Pietrini et al., 1998; Lauer et al., 1998). The digit span forward did not differ between patients and controls in one study (Lee et al., 1999), while it was reduced in two others compared to controls (Naëgelé et al., 1995) or norms (Verstraeten et al., 2000). Similarly, two studies (Knight et al., 1987; Lee et al., 1999) found no difference between patients and controls in the reversed digit span; another two (Naëgelé et al., 1995; Redline et al., 1997) found a reduced performance of patients, and a fifth study (Verstraeten et al., 2000) reported that some of the patients showed impaired performance in comparison to norms. The combined digit span did not differ between patients and controls in two studies (Knight et al., 1987; Lauer et al., 1998), while it was reduced in four studies compared to controls (Greenberg et al., 1987; Pietrini et al., 1998; Dani et al., 1996) or norms (Borak et al., 1996). In the visual domain, performance was reduced on the Corsi block-tapping task in one study (Naëgelé et al., 1995); on the Hiskey-Nebraska blocks, in one study (Pietrini et al., 1998) but not in another study (Lauer et al., 1998). In addition, Naëgelé et al. (1995) employed a double encoding task where a visual, a verbal and a double span were assessed, all of which were reduced in patients. Five studies (Naëgelé et al., 1995; Redline et al., 1997; Greenberg et al., 1987; Dani et al., 1996; Lee et al., 1999) reported means and standard deviations for attention span measures. The final data set compared 84 SRBD patients' performance to that of 71 controls on a combined digit span measure (Naëgelé et al., 1995; Greenberg et al., 1987; Dani et al., 1996; Lee et al., 1999) or the reversed digit span (Redline et al., 1997). Effect sizes ranged from -0.18 (Lee et al., 1999) to 2.30 (Dani et al., 1996) with significant between-study heterogeneity (x^2=9.61, df=4, p<0.05; Table 3). Figure 1 shows the individual study effect sizes.

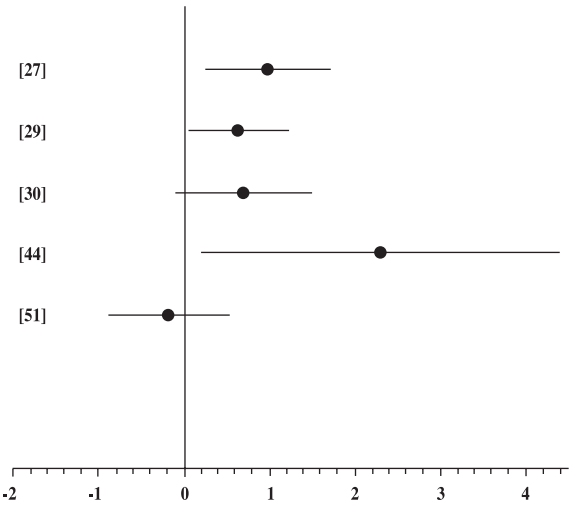

Figure 1: **Attention Span.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Focused attention was assessed in 22 studies (Findley et al., 1986, 1991; Sauter et al., 2000; Bédard et al., 1991; Stone et al., 1994; Borak et al., 1996; Kotterba et al., 1997, 1998; Cassel et al., 1989; Walsleben et al., 1989; Knight et al., 1987; Naëgelé et al., 1995; Redline et al., 1997: Greenberg et al., 1987; Kim et al., 1997; Phillips et al., 1994; Zozula et al., 1998a; Kuo et al., 2000; Lauer et al., 1998; Dinges et al., 1998; Verstraeten et al., 2000; Lee et al., 1999). In the majority of studies more than one test was employed. For this reason data will be first reviewed for separate tests and then for pooled measures of focused attention.

Nineteen studies compared Trail Making Test (TMT) performance of SRBD patients and control subjects (Bédard et al., 1991; Naëgelé et al., 1995, 1999; Redline et al., 1997; Greenberg at al., 1987; Kim et al., 1997; Phillips et al., 1994; Zozula et al., 1998a; Kuo et al., 2000; Findley et al., 1991; Lee et al., 1999; Kotterba et al., 1998) or population norms (Findley et al., 1986; Sauter et al., 2000, Kotterba et al., 1997, 1998; Cassel et al., 1989; Walsleben et al., 1989; Verstraeten et al., 2000). Thirteen studies employed the Trails A with seven of them (Redline et al., 1997; Phillips et al., 1994; Zolula et al., 1998a; Naëgelé et al., 1999; Kuo et al., 2000; Lauer et al., 1998; Lee et al., 1999) reporting no difference between patients and controls. Two studies (Naëgelé et al., 1995; Kotterba et al., 1998) found that patients scored lower than controls, and four studies (Kotterba et al., 1997; Cassel et al., 1989; Walsleben et al., 1989; Verstraeten et al., 2000) reported that patients' performance was impaired when compared to norms. Trails A effect sizes from only four studies were available (Naëgelé et al., 1995; Redline et al., 1997; Kuo et al., 2000; Lee et al., 1999), thus data integration was not undertaken. Fourteen studies used the Trails B with half of them (Naëgelé et al., 1995; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997; Kuo et al., 2000; Findley et al., 1991; Lee et al., 1999) finding no difference between patients and controls, while the others found some impairment in comparison to controls (Bédard et al., 1991; Zozula et al., 1998a; Naëgelé et al., 1999) or norms (Findley et al., 1986; Roehrs et al., 1995; Walsleben et al., 1989; Verstraeten et al., 2000). Eight studies (Bédard et al., 1991; Naëgelé et al., 1995; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997; Kuo et al., 2000; Findley et al., 1991; Lee et al., 1999) reported means and standard deviations and compared a total of 402 SRBD patients with 791 controls, with one study (Kim et al., 1997) contributing 199 patients and 642 controls. Effect sizes ranged from -0.02 (Lee et al., 1999) to 0.80 (Bédard et al., 1991) with no significant between-study heterogeneity (x^2=3.39, df=7, p>0.80). The pooled effect size for the fixed effects and the random effects model were identical and different from zero: $\Phi_{IV/DS}$=0.26 (SE($\Phi_{IV/DS}$)=0.06, z=3.96, p<0.01; Table 3, Figure 2). To examine the effect of study quality, we excluded the two studies (Bédard et al., 1991; Greenberg et al., 1987) that had the lowest level of evidence (Bédard et al., 1991). The remaining six effect sizes showed no significant between-study heterogeneity (x^2=1.49, df=5, p>0.90). The pooled effect size did not change substantially ($\Phi_{IV/DS}$=0.24 (SE($\Phi_{IV/DS}$)=0.07, z=3.62, p<0.01). On the Trails B performance of SRBD, patients showed a consistent but small reduction.

Table 3
Summary of the Meta-analytical Results

Neuropsychological Functions	Number of - Studies	Effect Sizes	SRBD Patients	Controls	Between-study Heterogeneity *Q*	d.f.	*p*	Random-effects Model Φ_{DS}	*SE*	*z*	*p*	Fixed-effects Model Φ_{IV}	*SE*	*z*	*p*
Perception	4				n.c.[b]			n.c.[b]				n.c.[b]			
Attention															
- Alertness	3				n.c.[b]			n.c.[b]				n.c.[b]			
- Attentional Span	5	5	84	71	9.61	4	<0.05	n.c.[a]				n.c.[a]			
- Focused Attention															
TMT, Trail A	4				n.c.[b]			n.c.[b]				n.c.[b]			
TMT, Trail B	8	8	402	791	3.39	7	0.84	0.26	0.06	3.96	<0.01	0.26	0.06	3.96	<0.01
SDST/DSST	4				n.c.[b]										
Cancellation Tests	5	5	280	702	5.03	4	0.28	0.35	0.24	2.89	<0.01	0.27	0.07	3.69	<0.01
Tests Pooled	9	9	420	802	2.92	8	0.94	0.27	0.06	4.19	<0.01	0.27	0.06	4.19	<0.01
- Divided Attention/Mental Tracking	6	6	158	93	6.17	5	0.28	0.23	0.15	1.50	ns	0.24	0.13	1.77	ns
- Selective Attention	5	4			n.c.[b]			n.c.[b]				n.c.[b]			
- Sustained Attention															
Reaction Times	6	6	254	229	8.38	5	0.13	0.31	0.13	2.37	<0.05	0.32	0.09	3.34	<0.01
Lapses	7	7	249	283	12.32	6	0.06	0.48	0.13	3.62	<0.01	0.36	0.09	3.99	<0.01
Quality of Performance	5	5	112	75	1.95	4	0.74	0.51	0.15	3.36	<0.01	0.51	0.15	3.36	<0.01
- Vigilance	2				n.c.[b]			n.c.[b]				n.c.[b]			
Motor Functions	5	5	312	764	25.08	4	<0.01	n.c.[a]				n.c.[a]			
Driving Simulation	9	9	398	331	6.36	8	0.61	0.61	0.08	7.78	<0.01	0.61	0.08	7.78	<0.01
Constructional Performance	2				n.c.[b]			n.c.[b]				n.c.[b]			
Memory															
- Immediate Recall	5	5	127	108	2.29	4	0.68	0.09	0.14	0.67	ns	0.09	0.14	0.67	ns
- Learning	2				n.c.[b]			n.c.[b]				n.c.[b]			
- Retrieval															
Visual Retrieval	5	5	109	74	4.30	4	0.37	0.67	0.17	4.00	<0.01	0.66	0.16	4.13	<0.01
Verbal Retrieval	7	7	340	736	6.25	6	0.39	0.17	0.08	2.16	<0.05	0.16	0.07	2.29	<0.05
- Complex/Integrated Measures	2				n.c.[b]			n.c.[b]				n.c.[b]			
Concept Formation, Reasoning, Executive Functions															
- Mental Flexibility	5	5	145	151	3.57	4	0.47	0.72	0.12	5.91	<0.01	0.72	0.12	5.91	<0.01
- Concept Formation	5	5	98	101	5.19	4	0.27	0.28	0.15	1.90	ns	0.31	0.17	1.82	ns
- Executive Functions					n.c.[b]			n.c.[b]				n.c.[b]			
Verbal Functions/Language Skills															
- Verbal Fluency	5	5	267	699	5.71	4	0.22	0.39	0.14	2.72	<0.01	0.25	0.07	3.37	<0.01
Composite Measures															
- General Intellectual Functioning	6	6	338	744	6.02	5	0.30	0.24	0.10	2.40	<0.05	0.21	0.07	3.01	<0.01

[a] n.c: not computed due to between-study heterogeneity; [b] n.c: not computed due to sparseness of data; TMT: Trial Making Test; DSST: Digit Symbol Substitution Test; SDST: Symbol Digit Substitution Test.

Symbol Digit Substitution Tests (SDST) and Digit Symbol Substitution Tests (DSST) were applied in eleven studies (Bédard et al., 1991; Stone et al., 1994; Roehrs et al., 1995; Borak et al., 1996; Walsleben et al., 1989; Redline et al., 1997; Kim et al., 1997; Phillips et al., 1994; Zozula et al., 1998a; Dinges et al., 1998; Verstraeten et al., 2000). Seven studies found no differences between SRBD patients and normal controls (Redline et al, 1997; Kim et al., 1997; Phillips et al., 1994; Zozula et al., 1998a; Dinges et al., 1998), insomniac controls (Stone et al., 1994), or norms (Stone et al., 1994; Roehrs et al., 1995). Four studies reported a reduced performance of patients compared with norms (Borak et al., 1996; Walsleben et al; 1989; Verstraeten et al., 2000) or controls (Bédard et al., 1991). No data integration was undertaken since only four studies provided means and standard deviations for the SDST (Stone et al., 1994; Redline et al., 1997; Kim et al., 1997) or the DSST (Bédard et al., 1991).

Ten studies investigated short-term attention with various cancellation tests, requiring the subjects to cancel either letters (Bédard et al., 1991; Knight et al., 1987; Redline et al., 1997; Greenberg et al., 1987; Lauer et al., 1998) or digits (Borak et al., 1996; Kotterba et al., 1997; Cassel et al., 1989; Naëgelé et al., 1995; Kim et al., 1997). Again, the results were mixed. While six studies found no difference between SRBD patients and normal controls (Kotterba et al., 1997; Knight et al., 1987; Redline et al., 1997; Kim et al., 1997; Lauer et al., 1987) or norms (Cassel et al., 1989), four others (Bédard et al., 1991; Borak et al., 1996; Kotterba et al., 1997; Greenberg et al., 1987) observed a reduced performance. Five studies reported means and standard deviations (Bédard et al., 1991; Naëgelé et al., 1995; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997) for a total of 280 SRBD patients and 702 control subjects. Effect sizes ranged from 0.18 (Redline et al., 1997) to 0.91 (Bédard et al., 1991) without significant between-study heterogeneity (x^2=5.03, df=4, p>0.20). The pooled effect size for the random effects model was Φ_{DS}=0.35 (SE(Φ_{DS})=0.24, z=2.89, p<0.01) and for the fixed effects model Φ_{IV}=0.27 (SE(Φ_{IV})=0.07, z=3.69, p<0.01; Table 3). The 95% confidence intervals of estimated population effect sizes did not include zero for either model (Figure 3). Performance in short-term cancellation tests showed a consistent small to moderate reduction in SRBD patients.

Pooled effect size estimations from the above analyses suggest that performance on different tasks of focused attention might be similar enough to reasonably combine them into an overall pooled effect size estimate. Nine studies provided means and standard deviations for TMT, DSST, SDST, or cancellation tests. If individual studies contributed more than one effect size, they were averaged to give a single effect size for each study. The nine studies were comprised of 420 SRBD patients and 802 control subjects. Within-study averaged effects sizes ranged from 0.08 (Redline et al., 1997) to 0.75 (Bédard et al., 1991) without significant between-study heterogeneity (x^2=2.92, df=8, p>0.90). The pooled effect size for the random effects and the fixed effects model were $\Phi_{IV/DS}$=0.27 (SE($\Phi_{IV/DS}$)=0.06, z=4.19, p<0.01; Table 3, Figure 4). The pooled effect size did not change substantially after excluding the two studies (Bédard et al., 1991; Greenberg et al., 1987) with the lowest level of evidence (Bédard et al., 1991): ($\Phi_{IV/DS}$=0.25 (SE($\Phi_{IV/DS}$)=0.07, z=3.78, p<0.01; x^2=1.09, df=6, p>0.90). In summary, SRBD patients experienced a small but very consistent reduction in short term focused attention.

Divided attention and mental tracking involve the ability to respond to more than one task at a time or to multiple elements or operations within a task (Lezak et al., 19985). This function was assessed in nine studies with a variety of tasks. In six studies, SRBD patients did not differ from controls in serial subtraction (Redline et al., 1997; Lee et al., 1999) serial

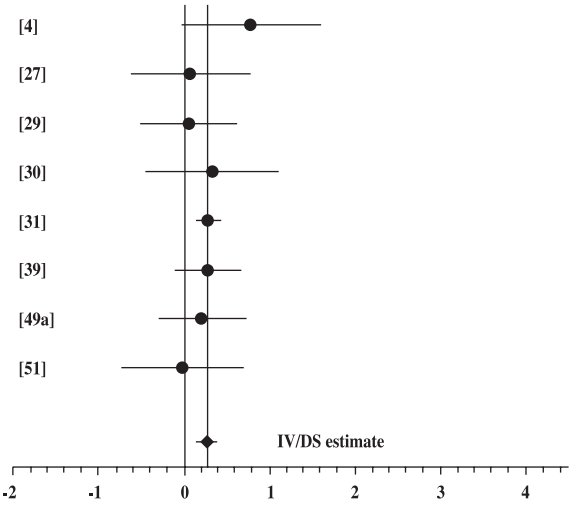

Figure 2: **Trail Making Test B.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

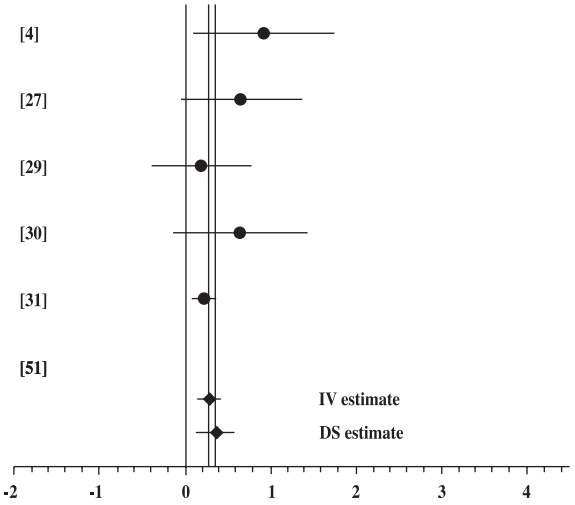

Figure 3: **Cancellation Tests.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

addition (Weeß et al., 1996; Rohmfeld et al., 1994), the N-back continuous attention memory task (Van Son et al., 2000), and overall performance in two German divided attention tasks (Weeß et al., 1996; Kotterba et al., 1998), although in one study (Kotterba et al., 1998) patients made more errors. For the Paced Auditory Serial Addition Test (PASAT) three studies reported a reduced performance of patients when compared to controls (Sloan et al., 1989; Findley et al., 1991) and norms (Findley et al., 1986). Finally, Sloan et al. (1989) reported a reduced performance of patients in a visual and an auditory tracking task. Six studies provided means and standard deviations for measures including the PASAT (Findley et al., 1991), serial subtraction (Redline et al., 1997; Lee et al., 1999), serial addition (Rohmfeld et al., 1994), a divided attention task (Rohmfeld et al., 1994), and a combined measure of divided attention and serial addition (Weeß et al., 1996). The studies compared the performance of a total of 158 SRBD patients and 93 controls. Effect sizes ranged from -0.21 (Weeß et al., 1996) to 0.66 (Findley et al., 1991) without significant between-study heterogeneity (x^2=6.17, df=5, p>0.20). The pooled effect size for the random effects model was Φ_{DS}=0.23 (SE(Φ_{DS})=0.15, z=1.50, n.s.) and for the fixed effects model Φ_{IV}=0.24 (SE(Φ_{IV})=0.13, z=1.77, n.s.; Table 3). The 95% confidence intervals of estimated population effect sizes did include zero for both models (Figure 5). Measures of divided attention and mental tracking showed no consistent reduction in SRBD patients.

Selective attention was assessed in nine studies (Verstraeten et al., 1996, 1997; Camus et al., 1999; Naëgelé et al., 1995; Phillips et al., 1994; Bonanni et al., 1999; Kuo et al., 2000; Lauer et al., 1987; Kotterba et al., 1998; Rohmfeld et al., 1994). SRBD patients showed a reduced performance in various tasks (Camus et al., 1999; Bonanni et al., 1999; Kotterba et al., 1998; Rohmfeld et al., 1994). The only task employed more than once was the Stroop Test. Here, three studies (Naëgelé et al., 1997; Kuo et al., 2000; Lauer et al., 1987) reported a reduced performance of patients when compared to healthy controls with only Philip et al. (Phillips et al., 1994) finding no difference for a group of elderly patients with very mild SRBD (mean AHI 11). Verstraeten et al. found that patients' performance on the Stroop Test did not differ from that of insomniac controls (Verstraeten et al., 1996) but was reduced when compared to a group of non-apneic snorers (Verstraeten et al., 1997). Only four independent studies (Verstraeten et al., 1996, 1997; Camus et al., 1999; Naëgelé et al., 1995; Kuo et al., 2000) provided means and standard deviations. For that reason no data integration was undertaken, although selective attention seems to be an area of possible impairment in SRBD patients.

Sustained attention was assessed in 14 studies with the Four Choice Reaction Time Test (FCRRT) (Bédard et al., 1991; Verstraeten et al., 1996, 1997; Zozula et al., 1998b; Morisson et al., 1997; Findley et al., 1991), the Psychometer Vigilance Test (PVT) (Muñoz et al., 2000; Barbé et al., 1998; Chugh et al., 1998; Dinges et al., 1998), the Continuous Performance Test (CPT) (Stone et al., 1994, Redline et al., 1997), and other tasks (Randerath et al., 2000; Kotterba et al., 1998; Rohmfeld et al., 1994). Reaction time was prolonged in patients compared to controls in five studies (Bédard et al., 1991; Muñoz et al., 2000; Barbé et al., 1998; Zozula et al., 1998b; Morisson et al., 1997) while in four other studies it did not differ from healthy controls (Findley et al., 1991), non-apneic snorers (Verstraeten et al., 1997), and norms (Findley et al., 1986; Stone et al., 1994). In comparison to insomniac controls, one study (Stone et al, 1994) found no difference while another (Verstraeten et al., 1996) reported that reaction times were faster for SRBD patients. Two studies found that patients made more errors than healthy controls (Morisson et al., 1997) and insomniacs (Verstraeten et al., 1996) while in another two there was no difference compared to controls (Bédard et al., 1991) and insomniacs (Stone et al, 1994). In six studies the number

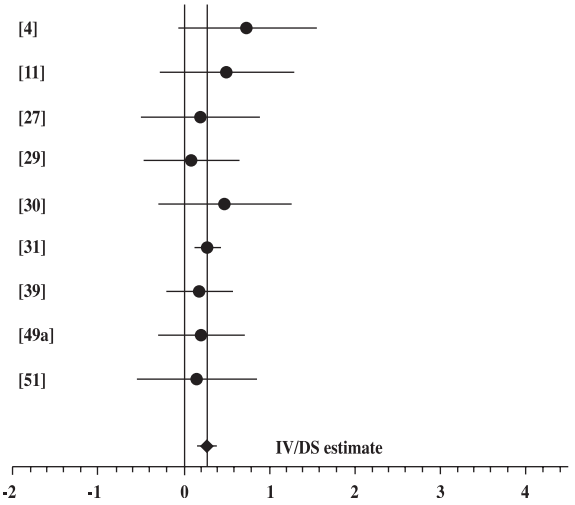

Figure 4: **Focused Attention.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

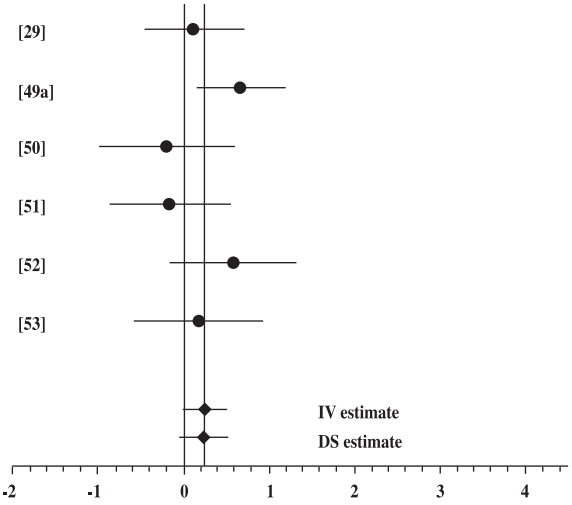

Figure 5: **Divided Attention and Mental Tracking.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

of gaps or lapses was higher in patients than controls (Bédard et al., 1991; Randerath et al., 2000; Zozula et al., 1998b; Dinges et al., 1998; Morisson et al., 1997), and non-apneic snorers (Chugh et al., 1998). Another study (Redline et al., 1997) found a reduced performance in patients–quantified by the signal detection measure d'–in the last but not the first two minutes of the ten-minute CPT. Two of three studies (Randerath et al., 2000; Kotterba et al., 1998; Rohmfeld et al., 1994) which used German sustained attention tasks reported a reduced performance of patients when compared to controls (Randerath et al., 2000; Kotterba et al., 1998). Twelve studies reported means and standard deviations. The different outcome measures (e.g., gaps, errors, reaction times) were grouped into three categories and analyzed separately. Seven studies reported mean reaction times for the CPT (Stone et al., 1994), FCRTT (Bédard et al., 1991; Verstraeten et al., 1996, 1997; Morisson et al., 1997; Findley et al., 1991), and the PVT (Muñoz et al., 2000; Barbé et al., 1998), comparing a total of 277 patients with 249 controls. One study (Morisson et al., 1997) reported means and standard deviations or standard errors that gave an effect size of 24.03 or 5.64, respectively. Since these effects sizes were at least five times larger than any of the other effect sizes, we decided to exclude this study from the following analysis. The combined sample size for the six studies was 254 patients and 229 controls. The remaining effect sizes ranged from -0.21 (Verstraeten et al., 1996, 1997) to 0.91 (Bédard et al, 1991) without significant between-study heterogeneity (x^2=8.38, df=5, p>0.10). The pooled effect size for the random effects model was Φ_{DS}=0.31 (SE(Φ_{DS})=0.13, z=2.37, p<0.05) and the fixed effects model was Φ_{IV}= 0.32 (SE(Φ_{IV})=0.09, z=3.34, p<0.01; Table 3). The 95% confidence intervals of estimated population effect sizes did not include zero for both models (Figure 5). For the second analysis, all outcome measures describing attentional lapses or time-on-task decrements were selected. Outcome measures included the number of gaps (FCRTT) in three studies (Bédard et al., 1991; Randerath et al., 2000; Morisson et al., 1997), number of lapses (PVT) (Chugh et al., 1998), transformed lapses (PVT) (Dinges et al., 1998), number of omission errors (CPT) (Stone et al., 1994), and a fatigue measure (PVT) (Barbé et al., 1998). The seven studies compared the performance of 249 SRBD patients and 273 controls. Effect sizes ranged from -0.19 (Stone et al., 1994) to 1.12 (Bédard et al., 1991) without significant between-study heterogeneity (x^2=12.32, df=6, p>0.05). The pooled effect size for the random effects model was Φ_{DS}=0.48 (SE(Φ_{DS})=0.13, z=3.62, p<0.01), and the fixed effects model was Φ_{IV}=0.36 (SE(Φ_{IV})=0.09, z=3.99, p<0.01; Table 3). The 95% confidence interval of estimated population effect sizes did not include zero for both models (Figure 6). In a final analysis, outcome measures describing quality of task performance were selected. This included percentage of errors (FCRTT) (Bédard et al., 1991; Morisson et al., 1997), number of errors (FCRTT) (Verstraeten et al., 1996, 1997), number of commission errors (CPT) (Stone et al., 1994), and the signal-detection parameter d' (CPT) (Redline et al., 1997). The five studies compared 112 patients and 75 controls. Effect sizes ranged from 0.14 (Stone et al., 1994) to 0.83 (Morisson et al., 1997) without significant between-study heterogeneity (x^2=1.95, df=4, p>0.70). The pooled effect size for the random effects model and the fixed effects model were identical and different from zero: $\Phi_{IV/DS}$=0.51 (SE($\Phi_{IV/DS}$)=0.15, z=3.36, p<0.01; Table 3, Figure 6). Sustained attention showed a consistent small to moderate reduction in SRBD patients. Measures describing the quality of performance (e.g. errors, gaps) seem to show a stronger effect than reaction time measures.

Vigilance performance was assessed in seven studies. SRBD patients performed poorer than controls in a variant of the Mackworth Clock Performance in one study (Weeß et al., 1996) but not different in another (Rohmfeld et al., 1998). Furthermore, Sauter et al., (2000), reported that approximately one third of the patients had scores below the 25th percentile although the average score of their SRBD patients was similar to the mean of a normal population. Kotterba and co-workers who tested vigilance in two studies (Kotterba et al., 1997, 1998), found that in one study (Kotterba et al., 1998) the average percentile rank of patients and controls did not differ while patients made a significantly higher number of mistakes. In the same study (Kotterba et al., 1998), none of the patients had a score below the 30th percentile, whereas in the other study (Kotterba et al., 1997), seven out of 40 SRBD patients had scores below the 25th percentile. Van Son et al. (Van Son et al., 2000) found no difference between SRBD patients and controls compared on the Parasumaran vigilance task. Finally, Schulz et al., (1997), used a modified CFF procedure repeatedly over a three-hour period and found a lower threshold in untreated, as compared to treated, patients. Since only two studies provided means and standard deviations (Schulz et al., 1997; Weeß et al., 1996) no data integration was undertaken. Vigilance performance in SRBD patients seems to be impaired in some, but not all, patients.

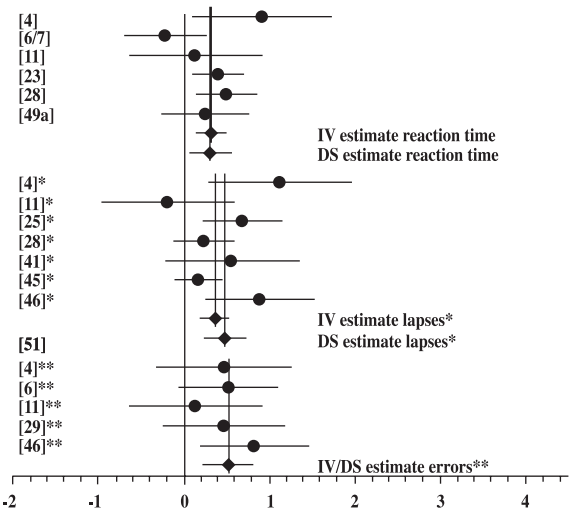

Figure 6: **Sustained Attention.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Motor Functions

Fifteen studies measured motor functions of the upper extremities. Finger tapping performance of SRBD patients did not differ from healthy controls (Knight et al., 1987; Phillips et al., 1994; Sloan et al., 1989), clinical controls (Stone et al., 1994; Verstraeten et al., 1996, 1997) and norms (Stone et al., 1994; Roehrs et al., 1995; Verstraeten et al., 2000) in eight studies. Performance on the grooved pegboard showed no difference between SRBD patients and controls in two studies (Kim et al., 1997; Kuo et al., 2000), while for the Purdue Pegboard, three studies found a reduced performance of SRBD patients when compared to controls (Bédard et al., 1991; Greenberg et al., 1987; Zozula et al., 1998a). Compared to population norms, one study (Walsleben et al., 1989) found no difference in patients, while another reported reduced performance (Verstraeten et al., 2000). Verstraeten et al. (1996, 1997) found no difference for non-dominant hand performance on the Purdue Pegboard for SRBD patients compared with insomniac patients and snorers, while performance of the dominant hand was reduced in patients when compared to snorers (Verstraeten et al., 1997) but not to insomniac patients (Verstraeten et al., 1996). Digit copying as employed by Stone et al. (1994), was not different for patients compared with insomniac controls or norms. Finally, Lee et al. (1999), found prolonged reaction times in a sensory motor task of SRBD patients compared to controls. Seven studies reported means and standard deviations for pegboard (Bédard et al., 1991; Verstraeten et al., 1996, 1997; Greenberg et al., 1987; Kim et al., 1997; Kuo et al., 2000) and tapping performance (Verstraeten et al., 1996, 1997), the digit-copying task (Stone et al., 1994), and the sensory motor task (Lee et al., 1999). In most of these studies performance of the dominant and non-dominant hand were given separately. Taken together, the studies compared 312 SRBD patients and 764 control subjects. All analyses including (i) averaging means and standard deviations across outcome measures within studies (seven effect sizes), (ii) selecting effect sizes from dominant hand performance only (six effect sizes), and (iii) restricting the analysis to the pegboard tasks (five effect sizes) showed significant between-study heterogeneity. Due to the small number of studies, further subdivision was not appropriate to reduce the remaining heterogeneity and the pooled effect sizes and confidence intervals were not computed. Figure 7 shows the individual study effect sizes for averaged performance measures, which are similar to the selected performance measures. Effect sizes ranged from 0.14 (Kim et al., 1997; Verstraeten et al., 1996, 1997) to 2.71 (Bédard et al., 1991) with between-study heterogeneity resulting mainly from study (Bédard et al., 1991). Single study evidence from the Level 1 study (Kim et al., 1997) suggests a smaller reduction in pegboard performance. Data integration was not undertaken due to significant heterogeneity and small number of studies.

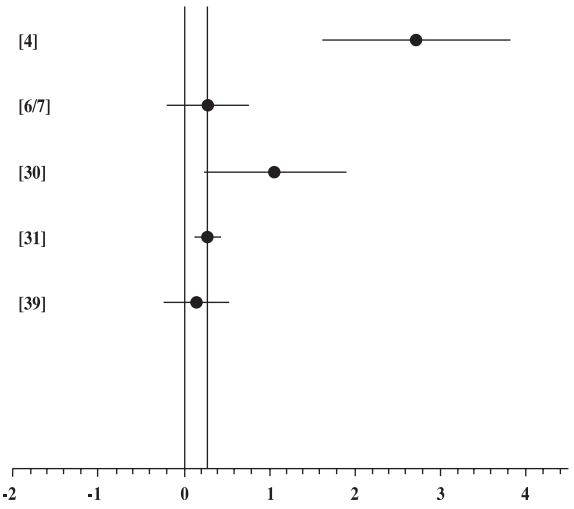

Figure 7: **Motor Functions.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Driving Simulation

Thirteen studies compared driving simulation performance of SRBD patients to healthy controls. With the exception of Ingram et al. (1994) and Findley et al. (1999), all other studies found reduced performance of patients (Findley et al., 1989, 1991, 1995; Risser et al., 2000; Muñoz et al., 2000; Juniper et al., 2000; George et al., 1996; Barbé et al., 1998; Dinges et al., 1998; Büttner et al., 2000). Twelve of the thirteen studies provided means and standard deviations and compared the performance of 507 SRBD patients to that of 509 control subjects. Since most studies reported more than one outcome measure, two separate analyses were undertaken. In the first analysis, performance measures were averaged within studies to yield a single effect size per study (average model). For the second analysis, we selected the outcome measure from each study that was most similar to the number or percentage of hits or steers (selection model). Both models exhibited significant heterogeneity (averaged model: x^2=22.99, df=11, p<0.02; selection model: x^2=21.94, df=11, p<0.03). To reduce heterogeneity, only those studies reporting number or percentage of hits, out-of-bound events, or steers passed were selected (Findley et al., 1989, 1991, 1995, 1999; Muñoz et al., 2000; George et al., 1996; Barbé et al., 1998; Ingram et al., 1994; Büttner et al., 2000). The analysis, which was based on ten effect sizes displayed significant heterogeneity (x^2=16.98, df=9, p<0.05). Visual inspection of the selected studies showed, that patients' average apnea severity measures clustered in the moderate to upper range from AHI 34 (Büttnet et al., 2000) to AHI 83 (Findley et al., 1989) with one study reporting an average AI of 19 (Ingram et al., 1994). After exclusion of this latter study, the analysis showed no statistically significant heterogeneity (x^2=6.36, df=8, p>0.60). The final data set thus included eight studies with nine effect sizes, comparing 398 SRBD patients and 331 controls. The pooled effect sizes and confidence intervals for the random effects model and the fixed effects model were identical: $\Phi_{IV/DS}$=0.61 (SE($\Phi_{IV/DS}$)=0.08, z=7.78, p<0.01; Table 3). The 95% confidence intervals for the estimates did not include zero

(Figure 8). Excluding the three studies with the lowest level of evidence (Findley et al., 1991, 1999; Büttner et al., 2000) did not change the pooled effect size substantially (random effects model: Φ_{DS}=0.58 (SE(Φ_{DS})=0.11, z=5.46, p<0.01; fixed effects model: Φ_{IV}=0.57 (SE(Φ_{IV})=0.10, z=5.72, p<0.01; x^2=5.44, df=5, p>0.30). In summary, there was a consistent moderate to large reduction in driving simulation performance of moderately to severely affected SRBD patients when compared to healthy controls.

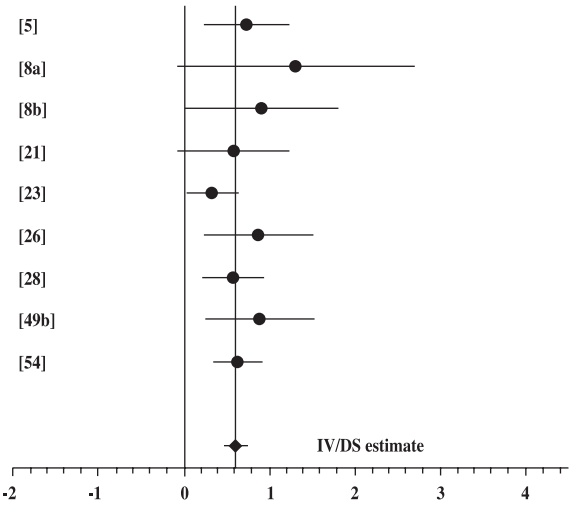

Figure 8: **Driving Simulation.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Constructional Performance

Eight studies compared constructional performance of SRBD patients with either healthy controls (Bédard et al., 1991; Knight et al., 1987; Greenberg et al., 1987; Zozula et al., 1998a; Pietrini et al., 1998) or population norms (Borak et al., 1996; Kales et al., 19885; Walsleben et al., 1989). Copying performance of SRBD patients was below that of controls in one study (Greenberg et al., 1987) and in a subgroup of more severely affected patients in another study (Bédard et al., 1991). Furthermore, Kales et al. (1985) found mild to severe impairment on the Bender Gestalt Test in twelve out of 50 patients, and signs of suspected impairment in another 26 in comparison to norms. While drawing performance in SRBD was reduced in one study (Pietrini et al., 1998), building and assembling performance did not differ between patients and controls (Knight et al., 1987; Greenberg et al., 1989; Zozula et al., 1998a) or norms (Walsleben et al., 1989) in four studies. Bédard et al. (1991) reported reduced building and assembling performance only for the subgroup of more severely affected patients, and Borak et al. (1996) found performance to be reduced when compared to population norms. Only two studies (Bédard et al., 1991; Greenberg et al., 1987) provided means and standard deviations of performance measures, so that no data integration could be undertaken.

Memory

There was a total of 27 studies on memory performance in SRBD patients, most of which reported multiple outcome measures (Findley et al., 1986, 1991; Lojander et al., 1999; Bédard et al., 1991; Verstraeten et al., 1996, 1997; Stone et al., 1994; Klonoff et al., 1987; Roehrs et al., 1995; Borak et al., 1996; Kales et al., 1985; Walsleben et al., 1989; Knight et al., 1987; Berry et al., 1990; Naëgelé et al., 1995, 1999; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997; Phillips et al., 1994; Zozula et al., 1998a; Sloan et al., 1989; Bonanni et al., 1999; Kuo et al., 2000; Pietrini et al., 1998; Lauer et al., 1998; Dinges et al., 1998; Lee et al., 1999; Kotterba et al., 1998).

Immediate recall for verbal material did not differ in six out of seven studies between SRBD patients and control subjects (Bédard et al., 1991; Knight et al., 1987; Greenberg et al., 1987; Findley et al., 1991; Lee et al., 1999) or norms (Findley et al., 1986). There was only one exception (Borak et al., 1996). For immediate visual recall, performance of SRBD patients was once again not different from that of healthy controls (Knight et al., 1987; Greenberg et al., 1987: Findley et al., 1991), clinical controls (Klonoff et al., 1987; Verstraeten 1996, 1997) or norms (Findley et al., 1986) in six out of nine studies, while three studies (Borak et al., 1996; Bédard et al., 1991; Pietrini et al., 1998) reported a reduced performance of SRBD patients. Six studies provided means and standard deviations for measures of immediate verbal or visual recall. There were four effect sizes for visual performance measures (Bédard et al., 1991; Verstraeten et al., 1996, 1997; Findley et al., 1989; Greenberg et al., 1987), four for verbal measures (Bédard et al., 1991; Findley et al., 1989; Greenberg et al., 1987; Lee et al., 1999), and three for both measures (Bédard et al., 1991; Findley et al., 1989; Greenberg et al., 1987). The final data set comprised five studies with five effect sizes for immediate memory performance, based on 127 SRBD patients and 108 control subjects. Effect sizes ranged from -0.21 (Verstraeten et al., 1996, 1997) to 0.29 (Bédard et al., 1991) without significant between-study heterogeneity (x^2=2.29, df=4, p>0.60). The pooled effect sizes for the random and the fixed effects model did not differ from each other and were not significantly different from zero ($\Phi_{IV/DS}$=0.09, SE($\Phi_{IV/DS}$)=0.14, z=0.67, n.s.; Table 3, Figure 9). Patients and healthy controls did not differ in measures of immediate recall.

Learning performance of SRBD patients was investigated in five studies. While verbal learning did not differ between patients and healthy controls in three of the studies (Knight et al., 1987; Redline et al., 1997; Kim et al., 1997), one study (Naëgelé et al., 1995) reported reduced performance in verbal and visual learning of SRBD patients in comparison to controls, and another (Borak et al., 1996) found learning performance on the AVLT "highly abnormal" in comparison to population norms. Since means and standard deviation were provided only for two studies (Naëgelé et al., 1995; Kim et al., 1997), the data could not be integrated.

Retrieval of verbal and visual material was investigated in 13 studies with measures of short (Stone et al., 1994; Knight et al., 1987) and long delay free recall (Findley et al., 1986, 1989;

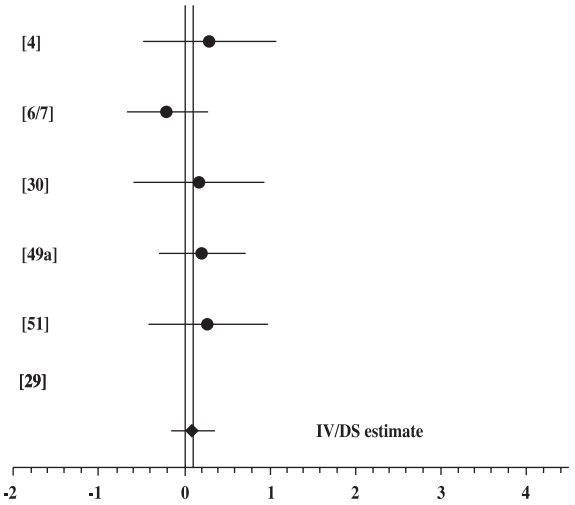

Figure 9: **Immediate Recall.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

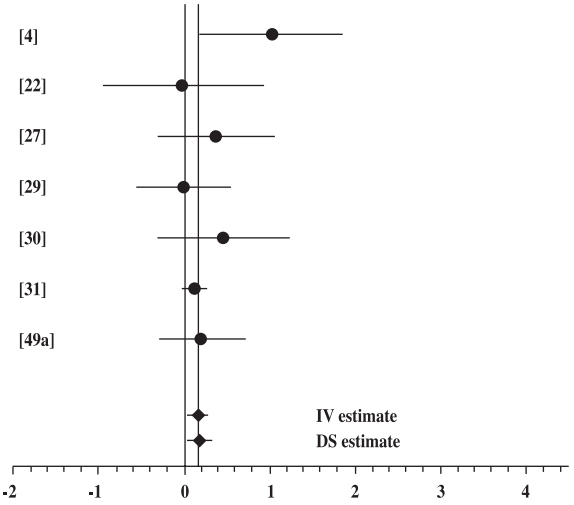

Figure 10: **Delayed Verbal Retrieval.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Bédard et al., 1991; Berry et al., 1990; Redline et al., 1997; Greenberg et al., 1987; Phillips et al., 1994; Zozula et al., 1998a; Pietrini et al., 1998), forgetting (Naëgelé et al., 1995; Greenberg et al., 1987; Kim et al., 1997) and recognition (Kim et al., 1997). Short-term retention with interference did not differ between SRBD patients and controls (Knight et al., 1987), insomniac controls (Stone et al., 1994) or norms (Stone et al., 1994). Similarly, measures of forgetting showed no difference between patients and controls in three studies (Naëgelé et al., 1995; Greenberg et al., 1987; Kim et al., 1997), as did recognition performance (Kim et al., 1997). Long delay free recall of visual material was assessed in eight studies with no difference in three studies (Greenberg et al., 1987; Phillips et al., 1994; Findley et al., 1991) and reduced performance of patients in three others (Berry et al., 1990; Zozula et al., 1998a; Pietrini et al., 1998). In a further study (Bédard et al., 1991), only a subgroup of severely but not mildly affected patients showed reduced visual recall, and finally patients with hypoxemia showed impairment in comparison to norms (Findley et al., 1986). Five studies (Bédard et al., 1991; Berry et al., 1990; Naëgelé et al., 1995; Greenberg et al., 1987; Findley et al., 1991) compared visual retrieval for a total of 109 SRBD patients and 74 healthy controls on measures of free recall (Bédard et al., 1991; Berry et al., 1990; Greenberg et al., 1987; Findley et al., 1991) or forgetting (Naëgelé et al., 1995). Effect sizes ranged from 0.40 (Findley et al., 1991) to 1.60 (Berry et al., 1990) without significant heterogeneity (x2=4.30, df=4, p>0.30). The pooled effect size was Φ_{DS}=0.67 (SE(Φ_{DS})=0.17, z=4.00, p<0.01) for the random effects model and Φ_{IV}=0.66 (SE(Φ_{IV})=0.16, z=4.13, p<0.01; Table 3) for the fixed effects model. The 95% confidence intervals did not include zero (Figure 10). Long delay free recall of verbal material was assessed in six studies. Patients and controls did not differ in four of them (Berry et al., 1990; Redline et al., 1997; Greenberg et al., 1987; Findley et al., 1991). In a fifth control group study only a subgroup of severely but not mildly affected patients showed reduced performance (Bédard et al., 1991), and in the sixth study patients group hypoxemia were impaired compared to norms (Findley et al., 1986). Seven studies (Findley et al., 1986; Bédard et al., 1991; Berry et al., 1990; Naëgelé et al., 1995; Redline et al., 1997; Greenberg et al., 1987; Kim et al., 1997) compared verbal retrieval of 340 SRBD patients and 736 healthy controls on measures of free recall (Findley et al., 1986, 1991; Bédard et al., 1991; Berry et al., 1990; Greenberg et al., 1986) or forgetting (Naëgelé et al., 1995, Kim et al., 1997). Effect sizes ranged from -0.02 (Berry et al., 1990) to 1.03 (Bédard et al., 1991) without significant between-study heterogeneity (x^2=6.25, df=6, p>0.30). The pooled effect size was Φ_{DS}=0.17 (SE(Φ_{DS})=0.08, z=2.16, p<0.05) for the random effects model and Φ_{IV}=0.16 (SE(Φ_{IV})=0.07, z=2.29, p<0.05; Table 3) for the fixed effects model. The 95% confidence intervals for both estimates did not include zero (Figure 11). Examining the effect of study quality on the stability of the pooled effect sizes was not feasible because that would have meant excluding three studies with the lowest evidence level (Bédard et al., 1991; Berry et al., 1990; Greenberg et al., 1986), which would have left only four studies in the analysis. The integrated analyses suggest that both delayed visual and verbal memory retrieval is impaired in patients with SRBD. While reductions were moderately to large for visual memory retrieval, they were small for verbal memory.

Complex or multiple integrated measures of memory performance were applied in 12 studies. These included the WMS logical and visual memory subscales (Roehrs et al., 1995; Phillips et al., 1994), the WMS full scale (Lojander et al., 1999; Walsleben et al., 1989; Dani et al., 1996), the CVLT full scale (Walsleben et al., 1989; Knight et al., 1987; Lauer et al., 1998) and other memory measures (Kales, et al., 1985; Kim et al., 1997; Sloan et al., 1989; Naëgelé et al., 1999; Kotterba et al., 1998). In most studies performance of SRBD patients was not different from that of healthy controls (Knight

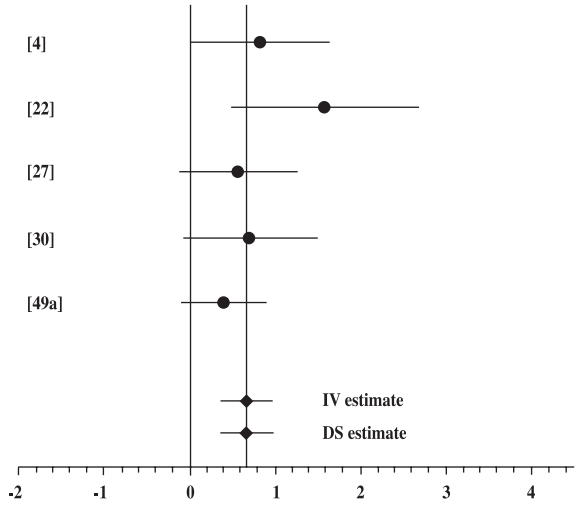

Figure 11: **Delayed Visual Retrieval.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

et al., 1987; Kim et al., 1997; Phillips et al., 1994; Naëgelé et al., 1999; Lauer et al., 1998; Kotterba et al., 1998) and population norms (Walsleben et al., 1989; Lojander et al., 1999). Only four studies found reduced memory performance in patients compared to controls (Sloan et al., 1989; Naëgelé et al., 1999; Dani et al., 1996) or norms (Kales et al., 1985). Yet, it should be noted that one quality Level 1 study (Kim et al., 1997) did not find any memory impairment in SRBD patients as compared to control subjects. Since means and standard deviations of performance measures were given in only two studies (Kim et al., 1997; Kotterba et al., 1998), no data integration was undertaken.

Other measures of memory performance included procedural learning (Redline et al., 1997; Naëgelé et al., 1999), working memory (Naëgelé et al., 1999; Kuo et al., 2000; Lee et al., 1999), incidental memory (Stone et al., 1994), a probed recall memory test (Dinges et al., 1998) and long-term memory impairment (Kales et al., 1985). Due to the diversity of performance measures no further data integration was undertaken.

In summary, there was no evidence for impaired immediate recall whereas delayed recall was either slightly (verbal memory) or moderately to severely (visual memory) impaired in SRBD patients.

Concept Formation, Reasoning and Executive Functions

There is little agreement in sleep-wake research (Camus et al., 1999) or neuropsychology (Miyake et al., 2000) about the exact nature of executive functions. Lezak's (1995) taxonomy distinguishes between concept formation, reasoning, and executive functions.

Concept formation denotes the ability to form concepts, to use categories, to generalize from single instances or to apply procedural rules and general principles (Lezak, 1995). Concept formation in verbal format has been assessed with the WAIS-R subtest Similarities in two studies (Bédard et al., 1991; Phillips et al., 1994), both of which found no difference between patients and controls. Concept formation in visual format was impaired in patients compared to healthy controls in one study (Lauer et al., 1998) and in comparison to norms (Roehrs et al., 1995), while three clinical control group studies found no difference between SRBD patients and insomniac (Verstraeten et al., 1996), snoring controls (Verstraeten et al., 1997) or control patients scheduled for surgery (Klonoff et al., 1987). Concept formation with special emphasis on sorting and shifting was assessed in four studies with the Wisconsin Card Sorting test (WCST) (Naëgelé et al., 1995, 1999; Redline et al., 1997; Lee et al., 1999). All four control group studies reported an increased number (Redline et al., 1997; Lee et al., 1999) or percentage (Naëgelé et al., 1995) of perseverative errors or a reduced overall WCST performance (Naëgelé et al., 1999). Other parameters like the number of errors (Naëgelé et al., 1995) or categories achieved (Naëgelé et al., 1995; Lee et al., 1999) did not differ between patients and controls.

Reasoning involves logical thinking, comprehension of relationships, and practical judgement (Lezak, 1995). Performance on the WAIS-R subtests Comprehension (Bédard et al., 1991), and Picture Completion (Borak et al., 1996), the 20-question procedure (Naëgelé et al., 1995) and a task requiring subjects to generate an "optimal" telegram did not differ between patients and controls (Bédard et al., 1991; Naëgelé et al., 1995), insomniac controls (Stone et al., 1994), or norms (Stone et al., 1994; Borak et al., 1996). On the WAIS-R subtests Picture Arrangement (Bédard et al., 1991; Borak et al., 1996) and Arithmetic (Borak et al., 1996) patients showed reduced performance in comparison to controls (Bédard et al., 1991) or norms (Borak et al., 1996).

Executive functions involve the four components volition, planning, purposive action, and effective performance (Lezak, 1995). Among these components planning was the focus of research in SRBD patients. Tower puzzles have been employed in two studies, one of which found no difference between patients and controls (Lee et al., 1999) while in the other patients showed reduced performance on the three- but not the four-disk task (Naëgelé et al., 1995). Maze tracing was reduced in patients in one study (Bédard et al., 1991) but did not differ between controls (Knight et al., 1987), insomniac controls (Stone et al., 1994) and norms (Stone et al., 1994) in two other studies. Another neuropsychological test that has been implicated (Miyake et al., 2000) as tapping executive functions is the Stroop test (Verstraeten et al., 1996, 1997; Naëgelé et al., 1995; Phillips et al., 1994; Kuo et al., 2000; Lauer et al., 1998), which was also included into the analysis.

Seven studies reported means and standard deviations of performance for the WCST (Naëgelé et al., 1995; Redline et al., 1997, Lee et al., 1999), the Stroop test (Verstraeten et al., 1996, 1997; Naëgelé et al., 1995; Kuo et al., 2000), Raven's progressive matrices (Verstraeten et al., 1996, 1997), various Tower tasks (Naëgelé et al., 1995; Lee et al., 1999), the 20-question task (Naëgelé et al., 1995), the telegram task (Stone et al., 1994) and of the WAIS-R subtests similarities,

comprehension, and picture arrangement (Bédard et al., 1991). Many studies employed more than one test and reported more than one outcome per test. To account for the multiplicity of outcome measures, stepwise analyses were undertaken. In a first step outcome-measures that assessed aspects of mental flexibility were selected, including the Stroop interference trial (Verstraeten et al., 1996, 1997; Naëgelé et al., 1995; Kuo et al., 2000) and perseverative errors in the WCST (Naëgelé et al., 1995; Redline et al., 1997; Lee et al., 1999). Five studies contributed five effect sizes, based on a total of 145 SRBD patients and 151 control subjects. No significant between-study heterogeneity was found (x^2=3.57, df=4, p>0.40), and the pooled effect sizes and confidence intervals for the random effects and fixed effects model were identical and different from zero, $\Phi_{IV/DS}$=0.72 (SE($\Phi_{IV/DS}$)=0.12, z=5.91, p<0.01; Table 3, Figure 12). SRBD patients showed a moderate to large reduction in mental flexibility.

The remaining outcome measures included the Tower tasks (Naëgelé et al., 1995; Lee et al., 1999), the 20-question task (Naëgelé et al., 1995), the telegram task (Stone et al., 1994), Raven's progressive matrices (Verstraeten et al., 1996, 1997), the number of categories achieved in the WCST (Naëgelé et al., 1995; Lee et al., 1999), errors in the WCST (Naëgelé et al., 1995), maze tracing (Bédard et al., 1991) and of the WAIS-R subtests similarities, comprehension, and picture arrangement (Bédard et al., 1991). Six studies contributed thirteen effect sizes. From these a subset of five studies was selected that assessed concept formation (Raven Progressive Matrices [Verstraeten et al., 1996, 1997], WAIS-R similarities [Bédard et al., 1991]), and reasoning performance (20-question task [Naëgelé et al., 1995], telegram task [Stone et al., 1994], WAIS-R comprehension and picture arrangement [Bédard et al., 1991]) as well as the sorting aspect of the WCST (number of categories achieved [Naëgelé et al., 1995; Lee et al., 1999]). Effect sizes were averaged so that five studies contributed five effect sizes comparing the performance of a total of 98 SRBD patients and 101 control subjects. No significant between-study heterogeneity was found (x^2=5.19, df=4, p>0.20). The pooled effect size for the random effects model was Φ_{DS}=0.31 (SE(Φ_{DS})=0.17, z=1.821, n.s.) and for the fixed effects model Φ_{IV}=0.28 (SE(Φ_{IV})=0.15, z=1.90, n.s.; Table 3). The 95% confidence intervals for both estimates did include zero (Figure 13). The remaining five effect sizes described aspects of executive functions but came from only three studies (Bédard et al., 1991; Naëgelé et al., 1995; Lee et al., 1999), which allowed no further analysis.

Although, there was a moderate to large reduction in mental flexibility aspects of performance (as indexed by the Stroop test and the WCST perseverative errors) of SRBD patients when compared with control subjects. No differences were found for aspects of concept formation and reasoning. Data on executive functions were not sufficient for quantitative synthesis.

Verbal Functions and Language Skills

Nine studies measured verbal fluency (Bédard et al., 1991), vocabulary (Knight et al., 1987; Greenberg et al., 1987), knowledge acquisition and retention (Borak et al., 1996; Walsleben et al., 1989; Greenberg et al., 1987; Ingram et al., 1994), and confrontation naming (Knight et al., 1987). Apart from verbal fluency, where one study (Bédard et al., 1991) reported a reduced performance in a subgroup of severely but not mildly affected SRBD patients in comparison to healthy controls, all other studies found no differences in performance of SRBD patients compared to control groups (Knight et al., 1987; Walsleben et al., 1989; Naëgelé et al., 1995; Greenberg et al., 1987; Kim et al., 1997; Lee et al., 1999) or population norms (Borak et al., 1996; Walsleben et al., 1989).

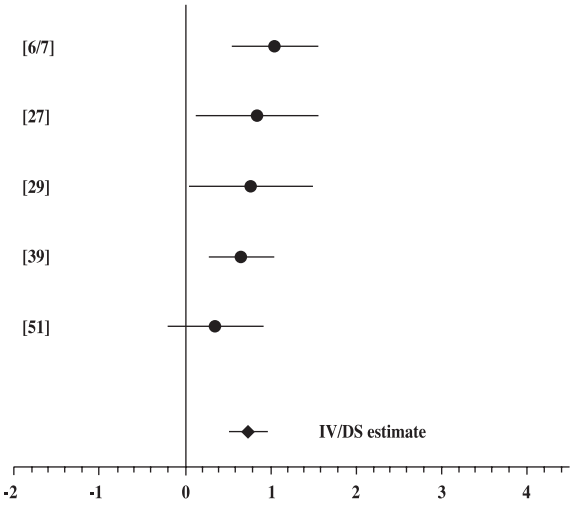

Figure 12: **Mental Flexibility.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

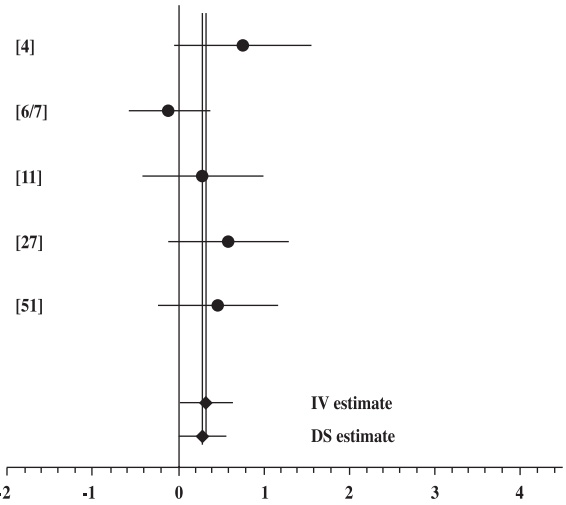

Figure 13: **Concept Formation, Sorting, and Reasoning.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Five studies reported means and standard deviations (or standard errors) for measures of verbal fluency (Bédard et al, 1991; Naëgelé et al., 1995; Greenberg et al., 1987; Kim et al., 1997; Lee et al., 1999) in 267 SRBD patients and 699 controls. Effect sizes ranged from 0.18 (Kim et al., 1997) to 0.72 (Bédard et al., 1991) without significant between-study heterogeneity (x^2=5.71, df=4, p>0.20). The pooled effect size for the random effects model was Φ_{DS}=0.39 (SE(Φ_{DS})=0.14, z=2.72, p<0.01) and for the fixed effects model Φ_{IV}=0.25 (SE(Φ_{IV})=0.07, z=3.37, p<0.01; Table 3). The 95% confidence intervals of estimated population effect sizes did not include zero for both models (Figure 14). The integrated data show a small to moderate but significant reduction in verbal fluency in SRBD patients compared to controls. For other measures of language skills, the number of studies was not sufficient for further integration.

Composite Measures

Composite measures are those measures that combine performances on widely different tasks into a single score. The best-known composite measures are the WAIS or WAIS-R verbal, performance or full scale IQ scores. While some (Lezak, 1995) maintain that such IQ-scores are neither useful nor informative in neuropsychological testing, others (Décary et al., 2000), within sleep research, have stressed the sensitivity of IQ-scores to hypoxemia.

General intellectual functioning of SRBD patients has been investigated with the full scale IQ of the WAIS-R (Bédard et al., 1991; Klonoff et al., 1987; Zozula et al., 1998a) and the WAIS (Borak et al., 1996; Pietrini et al., 1998) or its components verbal (WAIS-R [Bédard et al., 1991; Berry et al., 1987, 1990], WAIS [Lojander et al., 1999; Pietrini et al., 1998]) and performance IQ (WAIS-R [Bédard et al., 1991; Berry et al., 1987, 1990], WAIS [Lojander et al., 1999; Pietrini et al., 1998]). Some authors have estimated an IQ score by combining the vocabulary and block design subscales of the WAIS-R (Findley et al., 1986, 1991), or the Shipley Institute of Living Scale (Sloan et al., 1989). In older SRBD patients the Mini Mental Status Examination (MMSE) has been used to obtain a rough estimate of cognitive functioning (Berry et al., 1990; Phillips et al., 1994). Other measures that are presented here include a processing speed index (WAIS-III, [Pietrini et al., 1998]), a cognitive impairment evaluation, as indexed by a difference between WAIS verbal and performance IQ of greater than 15 points (Kales et al., 1985), and a factor-analytical derived measure of psychomotor efficiency (Kim et al., 1997) which included the Grooved pegboard, the symbol digit modalities test, the Trail B, a digit cancellation test, and the Controlled Word Association Test (COWAT).

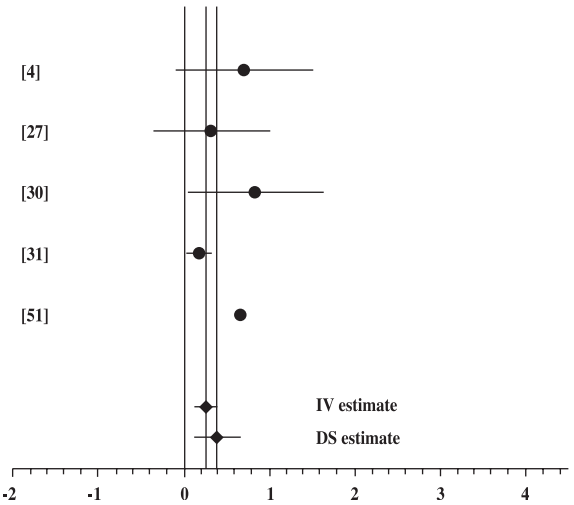

Figure 14: **Verbal Fluency.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

Six studies provided means and standard deviations of composite measures with three studies (Bédard et al., 1991; Berry et al., 1990; Pietrini et al., 1998) reporting more than one outcome measure. For the statistical analysis, those outcome measures were selected, which represented the most general intellectual functioning within each study, including the full scale IQ (WAIS-R [Bédard et al., 1991], WAIS [Pietrini et al., 1998]), the WAIS-R estimated IQ (Findley et al., 1991), the psychomotor efficiency factor (Kim et al., 1997), the processing speed index (Kuo et al., 2000) and the MMSE (Berry et al., 1990). Six studies contributed six effect sizes, based on a total of 338 SRBD patients and 744 control subjects. Effect sizes ranged from -0.10 (Berry et al., 1990) to 1.50 (Pietrini et al., 1998) with no significant between-study heterogeneity (x^2=6.02, df=5, p>0.30). The pooled effect size for the random effects model was Φ_{DS}=0.24 (SE(Φ_{DS})=0.10, z=2.40, p<0.05) and for the fixed effects model Φ_{IV}=0.21 (SE(Φ_{IV})=0.07, z=3.01, p<0.01; Table 3). The 95% confidence intervals for both estimates did not include zero (Figure 15).

Outcomes which were not included in the analysis were performance IQ (Bédard et al., 1991; Berry et al., 1990; Pietrini et al., 1998) and verbal IQ (Bédard et al., 1991; Berry et al., 1990; Pietrini et al., 1998). For verbal IQ scores, the majority of studies (Lojander et al., 1999; Bédard et al., 1991; Berry et al., 1987, 1990) found no reduction in verbal IQ score when compared to controls (Bédard et al., 1991; Berry et al., 1987, 1990) or population norms (Lojander et al., 1999) with one exception (Pietrini et al., 1998). The same five studies (Lojander et al., 1999; Bédard et al., 1991; Berry et al., 1987, 1990; Pietrini et al., 1998) also compared the performance IQ of SRBD patients either with controls subjects (Bédard et al., 1991; Berry et al., 1987, 1990; Pietrini et al., 1998) or population norms (Lojander et al., 1999). Two studies (Berry et al., 1990; Pietrini et al., 1998) found a reduced average performance IQ of patients. Furthermore, Bédard et al. (1991) reported a reduction in performance IQ for the subgroup of more severely affected patients (RDI>30) but not for the group of moderately affected patients (RDI<30). On the other hand, Lojander et al. (1999) found the median performance IQ of four subgroups of SRBD patients to be within the normal range, and Berry et al. (1987) reported no difference in performance IQ for an elderly group of patients when compared to controls.

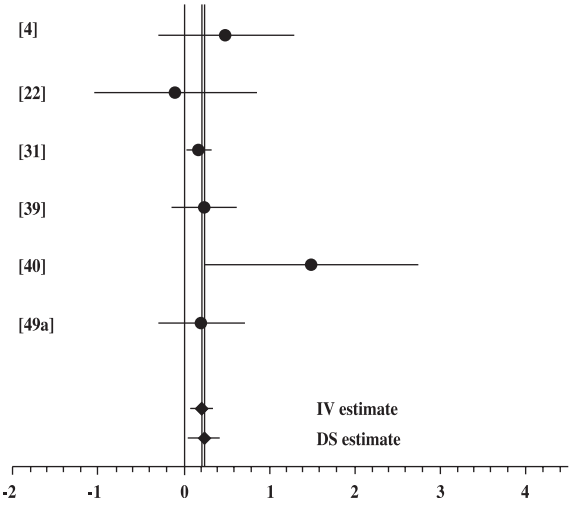

Figure 15: **General Intellectual Functioning.** Estimated effects estimations (g) and 95% confidence intervals for individual study outcomes, and pooled effect size estimations for the random-effects (DS) and inverse variance fixed-effects model (IV). The number in brackets on the left side gives the reference number of each study.

For measures of general intellectual functioning, a small but significant reduction in SRBD patients was observed when compared to control subjects. Additional comparisons for the performance IQ or the verbal IQ were not feasible due to the sparseness of data.

DISCUSSION

SRBD patients show a considerable impairment of performance over a wide range of neuropsychological functions when compared with healthy subjects, clinical control groups, or population norms. The main result of the meta-analysis was a moderate to large reduction in SRBD patients' performance in the areas of visual delayed memory retrieval, mental flexibility, and driving simulation. Small to moderate reductions of performance were found in the areas of focused and sustained attention, verbal delayed memory retrieval, verbal fluency, and composite measures of general intellectual functioning, while there were no differences for divided attention, concept formation and reasoning, and verbal or visual immediate memory performance. Due to the sparseness of studies or large between-study heterogeneity data integration was not possible in the areas of attentional span, motor functions, perception, alertness, selective attention, vigilance, constructional performance, learning performance, executive functions and verbal and performance IQ measures.

The 55 studies reviewed in the present paper cover a 15-year time span of growing research into cognitive performance of SRBD patients. There are several other reviews in this area (e.g., Hudgel, 1989; Kelly et al., 1990; Day et al., 1999; Décary et al., 2000; Engleman et al., 1999, 2000). The most recent ones are of particular interest since they were devoted specifically to neuropsychological functions of patients with SRBD (Décary et al., 2000; Engleman et al., 1999, 2000). Décary et al. (2000) have reviewed the main cognitive functions, suggested to be impaired in SRBD patients, and proposed a neuropsychological test battery for the assessment of cognitive deficits. They make several points that have a direct bearing on the present review. First, these authors reviewed basic neuropsychological functions (e.g., memory) and subfunctions (e.g., procedural memory) and found that single task performance, especially in the area of attention, has been subject to widely different interpretations from study to study. They stated that, "Such problems of interpretation will not be solved until neuropsychologists reach some form of agreement on the various concepts related to attention and how to measure it" (Décary et al., 2000, p. 371). In the present review, we have tried to solve this problem by grouping outcome measures as narrow as possible (e.g., TMT A and B separately) using the functional classification proposed by Lezak (1995). It must be acknowledged, however, that also this taxonomy is subject to discussion and has specific limitations, since it does, for example, not include computer-assisted tasks. The present results show that attentional performances of SRBD patients may vary for different tasks and even for different outcome parameters of the same task. Thus, the statistical aggregation of different outcome measures into overall effect size estimates relies on the assumption that the different measures are all operational realizations of the same underlying construct. The validity of such classifications should be examined in future research by systematic methods like multitrait-multimethod validation (Campbell and Fiske, 1959; see for an example, Strauss et al., 2000). Second, Décary et al. have identified several areas of cognitive functions where more research is required, namely short-term memory and procedural memory. The present review shows that this list is incomplete and other neuropsychological functions should be added: several attentional functions (attentional span, alertness, selective attention, vigilance), perception, motor and constructional performance, learning, executive functions, and composite measures. Third, and most importantly, this review has given much support to their proposal of a standard neuropsychological test battery for the assessment of cognitive deficits in SRBD patients. Tasks and outcome measures varied widely among studies, with the consequence that the profile of neuropsychological dysfunctions of SRBD patients remains incomplete although a substantial number of single studies was available.

Engleman et al. (1999, 2000) were the first to provide a quantitative overview of effect sizes for impairment in SRBD patients and to statistically integrate results across studies. Their data set consisted of case-control studies, comparing the performance of patients to that of healthy control groups. These authors have grouped outcome measures into broad neuropsychological functions such as attention and psychomotor tasks, memory and learning, executive and frontal lobe tasks. Their weighted average impairment effect sizes were 0.4 for attention, 0.2 for memory and 0.7 for executive and frontal scores. Although our findings are not directly comparable to theirs, due to very different aggregation levels, several points can be made. First, in the area of attention

and psychomotor performance, Engleman et al. (2000) have included attentional tasks as well as driving simulation performances, which have been treated separately in the present review. According to our analysis, driving simulation performance of SRBD patients seems to be more impaired (pooled effect size estimate 0.61) than attentional tasks (e.g., focused attention, effect size estimate 0.27). Second, for memory and learning performances, we again separated outcome measures for immediate memory, and verbal or visual delayed memory and found that performance of SRBD patients differs among these functional areas. Specifically, we found no difference between patients and controls in immediate memory performance, a small difference in verbal delayed memory, and a moderate to large difference in visual delayed retrieval. This suggests that different memory components must be distinguished to obtain an appropriate description of performance deficits of SRBD patients. Third, for executive and frontal lobe tasks, Engleman et al. (2000) included verbal fluency, mazes, and the WCST, among other tasks. They found a weighted average impairment effect size of 0.7, which is in agreement with the effect size 0.72 of the present review for mental flexibility as indexed by the WCST perseverative errors and the Stroop Interference Trial. We have, however, treated verbal fluency, different outcome parameters of the WCST, and mazes separately, and have therefore obtained results that differ in magnitude for our outcome aggregations. Executive functions in a more restricted sense, as defined by Lezak (1995), which included mazes and various tower tasks, were not integrated, due to the low number of studies. As discussed above, the grouping of neuropsychological tasks and the definition of levels of aggregation is an empirical question. From the present results we conclude that a detailed classification of neuropsychological tasks is needed for the attainment of an accurate pattern of neuropsychological dysfunction in SRBD patients.

Since the present findings of neuropsychological performance deficits are based on the application of meta-analytical techniques, peculiarities and limitations that are inherent to this technique must be considered (Glass et al., 1981). The "file drawer problem" and the integration of studies of different quality (Glass et al., 1981). The file drawer problem describes the fact that authors as well as journal editors tend to favor publications with statistically significant results, so that disproportional large effects may be over-represented in the published literature. The sole reliance on published evidence can thus seriously bias the conclusions drawn from a meta-analysis. The present review relied mainly on data published in journals, although we tried to limit this bias by integrating ongoing studies from conference proceedings. A likewise serious threat is the reliance on summary statistics as reported in the publications. Only 28 of a total of 44 studies provided these statistics. Those that did not report means and standard deviations often did not find any differences between patients and controls. What is desirable in such cases is to contact the authors and ask them for summary statistics and additional information not supplied in the published articles. This was not done due to time constraints for the present review, but is a much-needed step, although success of this procedure may be limited (Pilcher and Huffcutt, 1996, p. 320).

A meta-analysis is only as valid as the studies that contribute to it. If an analysis includes studies of low quality with systematic bias, this seriously threatens the validity of meta-analytical results. The present review included studies of very different methodological quality. Notably, there was only one study that could be evaluated as Level 1 evidence in analogy to established EBM levels of evidence. The majority of the studies clustered at Levels 3 and 4. Internal validity was generally undetermined, mainly due to the absence of precautions against selection or performance bias. It would be too optimistic a view to assume that such biases would be compensated through study-integration. It is far more likely that the bias is consistent across studies, especially because many studies relied on volunteer patients and control subjects. Patients who are willing to participate in a study might be those who show greater impairment and hope to profit from a thorough neuropsychological examination. Volunteer subjects, on the other hand, are known to exhibit superior performances and are willing to participate for this reason (Rosenthal and Rosnow, 1975). Comparing selected patients and controls might thus bias conclusions in the direction of larger effect sizes. An especially weak point of most studies was their lack of sampling. Only 21 studies drew actual samples from a predefined population (i.e., selection by means of a random mechanism). External validity related to case definition was generally high, and most studies followed standard diagnostic procedures. However, the diagnostic procedure was also a necessary inclusion criterion for the present review. This raises the question whether important evidence was overlooked that could have contributed to the review. In particular, two large-scale studies were excluded (Dealberto et al., 1996; Foley et al., 1999) where cases and controls were defined on the basis of participants' responses to questionnaire items. We maintain that self-report measures are not valid indicators for the presence, and more importantly, the absence of SRBD. We have used five quality criteria for the evaluation of non-interventional studies. It remains to be determined whether they are sufficient for the evaluation of these types of studies. Furthermore, quality evaluation was done by only one of the authors so that reproducibility was not checked. The relationship of quality criteria to study outcomes of meta-analyses is not well understood (Clarke and Oxman, 2000). In general (Clarke and Oxman, 2000; Hunter and Schmidt, 1990; Lau et al., 1998) exploring the effects of differing quality by including or excluding studies, or by meta-regressional approaches is recommended so that the influence of low quality studies can be assessed statistically. This was done in the case of the TMT B, focussed attention, and driving simulation without substantial changes in effect sizes. In the majority of the analyses exclusion of studies was not feasible because the number of studies within each analysis was generally small. It should also be mentioned that quality

ratings were not independent from other characteristics which are likely to influence effect sizes. In particular, the only Level 1 study used a population based approach and for that reason the group of participants with SRBD exhibited a low average apnea index which in turn seems likely to influence cognitive performance. More studies or more statistical information from the available studies are needed to meaningfully explore the influence of study quality, apnea severity or other variables on cognitive performance in SRBD patients.

One final point concerns the accuracy of the present results. Although the analysis yielded a reduced performance of SRBD patients when compared with control subjects, one might find the magnitude of these effects disappointingly small when compared to single study outcomes or compared to the effects of sleep deprivation (Pilcher and Huffcutt, 1996). Apart from the considerations detailed above, there are several further reasons for the lower limits of performance differences between SRBD patients and controls in the present review. First, we did not attempt to correct effect size estimations for varying reliability of data (Hunter and Schmidt, 1990). Indeed, the American Psychological Association (APA) states that interpreting the size of observed effects requires the assessment of the reliability of the scores, since attenuation by reliability can be substantial (Wilkinson, 1999). We did not incorporate reliability in our analysis because of the universal lack of reliability estimates of SRBD patients' performance on neuropsychological tasks. Furthermore, even where such data is available (for example, Ingram et al., 1999), reliability computations suggest large differences among patient groups and healthy controls. It is to be expected, however, that a correction for reliability will yield substantially larger effect sizes. Second, it was the objective of the present study to review cognitive dysfunctions in SRBD patients. As such, the study populations included varied widely with regard to age, apnea severity and other aspects. We have found significant between-study heterogeneity for attentional span measures and motor functions. Other areas of neuropsychological functions showed more homogeneous effect sizes, but tests for heterogeneity possess only low statistical power (Lau et al., 1998). Thus, we may have pooled effect sizes from different studies which might have been better described by separate effect size estimations. However, the small number of studies made further subdivisions of data sets infeasible. Another important point is that single studies with a heterogeneous sample of patients, due to differences in age or apnea severity, will attenuate effect sizes. Studies that more narrowly define the group of patients with regard to parameters that are expected to correlate with cognitive performance measures will likely yield larger effect sizes solely by decreasing within-group variability. Thus, it is to be expected that, as primary and secondary research designs become more focused, the results will differ from those of the present review.

SRBD patients show cognitive dysfunction in many areas of neuropsychological task performance. The pattern of this dysfunction, however, is not a simple one. Our results show that the magnitude of cognitive dysfunction varies for different areas of neuropsychological performance and within basic functions and subfunctions for different task parameters. How these are related to cognitive function in the real world is unknown. Further research is needed to clarify which factors cause the cognitive dysfunction in SRBD patients, what the consequences of this dysfunction are, and whether treatment of SRBD is effective in reducing cognitive dysfunction.

ACKNOWLEDGMENTS

The authors do not wish to include any acknowledgements.

REFERENCES

Study References

1. Barbé F, Pericás J, Muñoz A, Findley L, Antó JM, Agusti AG. Automobile accidents in patients with sleep apnea syndrome. An epidemiological and mechanistic study. *Am J Respir Crit Care Med* 1998; 158: 18-22.
2. Bédard MA, Montplaisir J, Richer F, Rouleau I, Malo J. Obstructive sleep apnea syndrome: Pathogenesis of neuropsychological deficits. *J Clin Exp Neuropsychol* 1991; 13: 950-64.
3. Berry DT, Phillips BA, Cook YR, Schmitt FA, Gilmore RL, Patel R, Keener TM, Tyre E. Sleep-disordered breathing in healthy aged persons: Possible daytime sequelae. *J Gerontol* 1987; 42: 620-6
4. Berry DTR, Phillips BA, Cook YR, Schmitt FA, Honeycutt NA, Arita AA, Allen RS. Geriatric sleep apnea syndrome: a preliminary description. *J Gerontol* 1990; 45: M169-74.
5. Bonanni E, Gori S, Belloli S, Pasquali L, Maestri M, Berretini S, De Vito A, Murri L. Neuropsychological evaluation of vigilance in patients with obstructive sleep apnea before and after L-UPP. *Sleep Res Online* 1999; 2(Suppl. 1): 332.
6. Borak J, Cieslicki JK, Koziej M, Matuszewski A, Zieli_ski J. Effects of CPAP treatment on psychological status in patients with severe obstructive sleep apnea. *J Sleep Res* 1996; 5: 123-7.
7. Büttner A, Siller K, Kraft-Malycha A, Randerath W, Rühle KH. Normwerte und Gütekriterien eines interaktiven Fahrsimulators ("carsim"). *Somnologie* 2000; 4: 129-36.
8. Camus JF, Bacqué MF, Fleury B. Evaluation of attentional processing impairment in obstructive sleep apnea syndrome by means of a dichotic listening task. Cahiers de *Psychologie Cognitive/ Current Psychology of Cognition* 1999; 18: 45-73.
9. Cassel W, Stephan S, Ploch T, Peter JH. Psychologische Aspekte schlafbezogener Atemregulationsstörungen. *Pneumologie* 1989; 43: 625-9.
10. Chugh DK, Weaver TE, Dinges DF. Psychomotor vigilance performance in sleep apnea patients compared to patients presenting with snoring without sleep apnea. *Sleep* 1998; 21(Suppl. 3): 159.

11. Dani A, Pietrini P, Furey M, Freo U, Raphaelson M, Alexander GE, Horwitz B, Guazzelli M, Gemignani A. Shapiro MC. Patients with sleep apnea syndrome (SAS) show neuropsychological impairment and regional cerebral glucose metabolic deficits: A pre-treatment positron emission tomography (PET) study. *J Sleep Res* 1996; 5(Suppl. 1): 43.
12. Dinges DF, Maislin G, Staley B, Pack F, Woodle C, Pack A. Sleepiness and neurobehavioral functioning in relation to apnea severity in a cohort of commercial motor vehicle operators. *Sleep* 1998; 21(Suppl. 3): 83.
13. Findley LJ, Barth JT, Powers DC, Wilhoit SC, Boyd DG, Suratt PM. Cognitive impairment in patients with obstructive sleep apnea and associated hypoxemia. *Chest* 1986; 90: 686-90.
14. Findley LJ, Fabrizio MJ, Knight H, Norcross BB, Laforte AJ, Suratt PM. Driving simulator performance in patients with sleep apnea. *Am Rev Respir Dis* 1989; 140: 529-30.
15. Findley LJ, Presty SK, Barth JT, Suratt PM. Impaired cognition and vigilance in elderly subjects with sleep apnea. In: ST Kuna, PM Suratt, JE Remmers (Eds). *Sleep and respiration in the aging adults*. Elsevier, New York 1991: 259-63.
16. Findley L, Unverzagt M, Guchu R, Fabrizio M, Buckner J Suratt P. Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. *Chest* 1995; 108: 619-24.
17. Findley LJ, Suratt PM, Dinges DF. Time-on-task decrements in "Steer Clear" performance of patients with sleep apnea and narcolepsy. *Sleep* 1999; 22: 80[illegible]-9
18. George CF, Bourdeau AC, Smiley A. Comparison of simulated driving performance in narcolepsy and sleep apnea patients. *Sleep* 1996; 19: 713-17.
19. Greenberg GD, Watson RK, Deptula D. Neuropsychological dysfunction in sleep apnea. *Sleep* 1987; 10: 254-62.
20. Ingram F, Henke KG, Levin HS, Fishel Ingram PT, Kuna ST. Sleep apnea and vigilance performance in a community-dwelling older sample. *Sleep* 1994; 17: 248-52.
21. Juniper M, Hack MA, George CF, Davies RJ, Stradling JR. Steering simulation performance in patients with obstructive sleep apnea and matched control subjects. *Eur Respir J* 2000; 15: 590-5.
22. Kales A, Caldwell AB, Cadieux RJ, Vela-Bueno A, Ruch LG, Mayes S. Severe obstructive sleep apnea – II: associated psychopathology and psychosocial consequences. *J Chronic Dis* 1985; 38: 427-34.
23. Kim HC, Young T, Matthews CG, Weber SM, Woodward AR, Palta M. Sleep-disorderd breathing and neuropsychological deficits. *Am J Respir Crit Care Med* 1997; 156: 1813-19.
24. Klonoff H, Fleetham J, Taylor DR, Clark C. Treatment outcome of obstructive sleep apnea. Physiological and neuropsychological concomitants. *J Nerv Ment Dis* 1987; 175: 208-12.
25. Knight H, Millman RP, Gur RC, Saykin AJ, Doherty JU, Pack AI. Clinical significance of sleep apnea in the elderly. *Am Rev Respir Dis* 1987; 136: 845-50.
27. Kotterba S, Widdig W, Duscha C, Rasche K. Ereigniskorrelierte Potentiale und neuropychologische Untersuchung bei Schlafapnoepatienten. *Pneumologie* 1997; 51: 712-15.
26. Kotterba S, Rasche K, Widdig W, Blombach S, Duchna K, Duchna HW, Schultze-Werninghaus G, Malin JP. Vigilance and neuropsychological capacity in obstructive sleep apnea syndrome and chronic obstructive pulmonary disease. *Somnologie* 1998; 2: 117-22.
28. Kuo TF, Bootzin RR, Quan SF, Hiley A, Caccappolo E, Walsleben JA, Kaszniak AW, Reminger SL. Sleep-disordered breathing and neuropsychological functioning: A study in non-patient adults age 45-75. *Sleep* 2000; 23 (Suppl. 2): 58.
29. Lauer CJ, Fischer J, Lund R, Zihl J. Patients with obstructive sleep apnea syndrome (OSAS) show deficits in attentional demands and problem solving capacities. *J Sleep Res* 1993; 76 (suppl. 2): 148.
30. Lee MM, Strauss ME, Adams N, Redline S. Executive functions in persons with sleep apnea. *Sleep and Breathing* 1999; 3: 13-16.
31. Lojander L, Kajaste S, Maasilta P, Partinen M. Cognitive function and treatment of obstructive sleep apnea syndrome. *J Sleep Res* 1999; 8: 71-6.
32. Morisson F, Decary A, Poirier G, Godbout R, Bédard M-A, Lavigne G, Montplaisir J. Vigilance impairment in obstructive sleep apnea syndrome, idiopathic hypersomnia and narcolepsy. *Sleep Res* 1997; 26: 438.
33. Muñoz A, Mayoralas LR, Barbé F, Péricás J, Austi´s AG. Long-term effects of CPAP on daytime functioning in patients with sleep apnea syndrome. *Eur Respir J* 2000; 15: 676-81.
34. Naëgelé B, Thouvard V, Pépin JL, Lévy P, Bonnet C, Perret JE, Pellat J, Feuerstein C. Deficits of cognitive executive functions in patients with sleep apnea syndrome. *Sleep* 1995; 18: 43-52.
35. Naëgelé B, Mazza S, Pépin JL, Lévy P, Feuerstein C. Does sleep fragmentation associated with obstructive respiratory events cause neuropsychological deficits? *Sleep Res Online* 1999; 2(Suppl. 1): 409.
36. Phillips BA, Berry DT, Schmitt FA, Harbinson L , Lipke-Molby T. Sleep-disordered breathing in healthy aged persons: two- and three-year follow up. *Sleep* 1994; 17: 411-15.
37. Pietrini P, Dani A, Raphaelson M, Furey ML, Alexander G, Levine B, Guazzelli M, Rapoport S, Shapiro MC. Cerebral glucose metabolic and neuropsychological dysfunction in patients with untreated sleep apnea syndrome (SAS). *Sleep* 1998; 21(Suppl. 3): 82.
38. Randerath WJ, Gerdesmeyer C, Siller K, Gil G, Sanner B, Rühle K-H. A test for the determination of sustained attention in patients with obstructive sleep apnea syndrome. *Respiration* 2000; 67: 526-32.

39. Redline S, Strauss ME, Adams N, Winters M, Roebuck T, Spry K, Rosenberg C, Adams K. Neuropsychological function in mild sleep-disordered breathing. *Sleep* 1997; 20: 160-7.
40. Risser MR, Ware JC, Freeman FG. Driving simulation with EEG monitoring in normal and obstructive sleep apnea patients. *Sleep* 2000; 23: 393-8.
41. Roehrs T, Merrion M, Pedrosi B, Stepanski E, Zorick F, Roth T. Neuropsychological function in obstructive sleep apnea syndrome (OSAS) compared to chronic obstructive pulmonary disease (COPD). Sleep 1995; 18: 382-8.
42. Rohmfeld R, Weeß HG, Föller-Schlums G, Schneider C, Meyer J, Steinberg R, Pritzel M. Beeinträchtigung der Aufmerksamkeit und Vigilanz bei Patienten mit obstruktivem Schlaf-Apnoe-Syndrom. *Wiener Medizinische Wochenschrift* 1994; 144: 74-8.
43. Sauter C, Asenbaum S, Popovic R, Bauer H, Lamm C, Klösch G and Zeitlhofer J. Excessive daytime sleepiness in patients suffering from different levels of obstructive sleep apnea syndrome. *J Sleep Res* 2000; 9: 293-301.
44. Schulz H, Wilde-Frenz J, Grabietz-Kurfürst U. Cognitive deficits in patients with daytime sleepiness. *Acta Neurol Belg* 1997; 97: 108-12.
45. Sloan K, Craft S, Walsh JK. Neuropsychological function in obstructive sleep apnea with and without hypoxemia. *Sleep Res* 1989; 18: 304.
46. Stone J, Morin CM, Hart RP, Remsberg S, Mercer J. Neuropsychological functioning in older insomniacs with or without obstructive sleep apnea. *Psychol Aging* 1994; 2: 231-6.
47. Van Son B, Hofman WF, van Uffelen K. Sleep fragmentation, as measured by EEG and PTT arousals, and daytime functioning in OSA patients. *J Sleep Res* 2000; 9(Suppl. 1): 199.
48. Verstraeten E, Cluydts R, Verbraecken J, de Roeck J. Neuropsychological functioning and determinants of morning alertness in patients with obstructive sleep apnea syndrome. *J Internat Neuropsychol Soc* 1996; 2: 306-14.
49. Verstraeten E, Cluydts R, Verbraecken J, de Roeck J. Psychomotor and cognitive performance in nonapneic snorers: preliminary findings. *Percept Mot Skills* 1997; 84: 1211-22.
50. Verstraeten E, Hoffmann G, Cluydts R. Attentional capacity and/or executive control deficits in obstructive sleep apnea. *J Sleep Res* 2000; 9(Suppl. 1): 201.
51. Walsleben JA, Squires NK, Rothenberger VL. Auditory event-related potentials and brain dysfunction in sleep apnea. *Electroencephal Clin Neurophysiol* 1989; 74: 297-311.
52. Weeß HG. *Leistungserfassung beim obstruktiven Schlafapnoesyndrom*. Regensburg: S. Roderer, 1996.
53. Zozula R, Caccappolo E, Rapoport D, Walsleben J. Assessment of cognitive function in moderate-severe OSAS. *Sleep* 1998a; 21(Suppl. 3): 71.
54. Zozula R, Walsleben JA, Rapoport DM, Campbell SS, Spielman AJ. Discordant objective daytime performance measures following CPAP treatment for OSAS. *Sleep* 1998b; 21(Suppl. 3): 155.

Non-study References

55. Campbell GT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 1959; 56: 81-105.
56. Clarke M, Oxman AD, eds. *Cochrane Reviewers´ Handbook 4.1* [updated June 2000]. In: Review Manager (RevMan) [Computer program]. Version 4.1. Oxford, England: The Cochrane Collaboration, 2000.
57. Cohen, J. A power primer. *Psychological Bulletin* 1992, 112: 115-159.
58. Day R, Gerhardstein R, Lumley A, Roth T, Rosenthal L. The behavioral morbidity of obstructive sleep apnea. *Progress in Cardiovascular Diseases* 1999; 41: 341-354.
59. Dealberto M-J, Pujol F, Courbon D, Alpérovitch A. Breathing disorders during sleep and cognitive performance in an older community sample: The EVA study. *J Am Geriatr Soc* 1996, 44: 1147-1294.
60. Décary A, Rouleau I and Montplaisir J. Cognitive deficits associated with sleep apnea syndrome: A proposed neuropsychological test battery. *Sleep* 2000; 23: 369-381.
61. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; 7: 177-188.
62. Diagnostic Classification Steering Committee, Thorpy MJ, Chairman. *International classification of sleep disorders: Diagnostic and coding manual*. Rochester, Minnesota: American Sleep Disorders Association 1990
63. Engleman H, Joffe D. Neuropsychological function in obstructive sleep apnea. *Sleep Medicine Reviews* 1999; 3: 59-78.
64. Engleman HM, Kingshott RN, Martin SE, Douglas NJ. Cognitive function in sleep apnea/hypopnea syndrome (SAHS). *Sleep* 2000; 23 (Suppl. 4): 102-108.
65. Foley DJ, Monjan AA, Masaki KH, Enright PL, Quan SF, White LR. Association of symptoms of sleep apnea with cardiovascular disease, cognitive impairment, and mortality among older Japanese-American men. *J Am Geriatr Soc* 1999, 47: 524-514.
66. Glass GV, McGaw B, Smith ML. *Meta-analysis in social research*. Sage: Beverly Hills, CA, 1981.
67. Hedges LV, Olkin I. *Statistical methods for meta-analysis*. New York: Academic Press, 1985.
68. Hudgel DW. Neuropsychiatric manifestations of obstructive sleep apnea: A review. *Int J Psychol Med* 1989; 19: 11-22.
69. Hunter JE, Schmidt FL. *Methods of meta-analysis: Correcting error and bias in research findings*. Sage: Newbury Park, CA, 1990.
70. Ingram F, Greve KW, Ingram Fishel PT, Soukup VM. Temporal stability of the Wisconsin Card Sorting Test in an untreated patient sample. *British J of Clin Psychol* 1999; 38: 209-211.

71. Johnson, BT, Eagly AH. *Quantitative synthesis of social psychological research*. In: HT Reis, CM Judd (Eds). Handbook of Research Methods in Social and Personality Psychology. Cambridge University Press, Cambridge 2000: 496-528.
72. Kelly DA, Claypole KH, Coppel DB. Sleep apnea syndrome: symptomatology, associated features and neurocognitive correlates. *Neuropsychology Review* 1990; 1: 323-342.
73. Lau J, Ioannidis JP, Schmidt CH. *Summing up evidence: one answer is not always enough*. Lancet 1998; 351: 123-127.
74. Lezak MD. *Neuropsychological assessment*. 3rd ed. Oxford University Press: New York, 1995.
75. Miyake A, Friedman NP, Emerson MJ, Witzki AH and Howerter A. The unity and diversity of executive functions and their contribution to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychol* 2000; 41: 49-100.
76. Pilcher JJ, Huffcutt AI. Effects of sleep deprivation on performance: a meta-analysis. *Sleep* 1996; 19: 318-326.
77. Poole C, Greenland S. Random-effects meta-analysis are not always conservative. *Am J Epidemiol* 1999, 150: 469-475.
78. Rosenthal R, Rosnow RL. *The volunteer subject*. Wiley, New York, 1975.
79. Strauss ME, Thompson P, Adams NL, Redline S, Burant C. Evaluation of a model of attention with confirmatory factor analysis. *Neuropsychol* 2000; 14: 201-208.
80. Voyer D, Voyer S, Bryden MP. Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin* 1995, 117: 250-270.
81. Wilkinson L, The APA Task Force on Statistical Interference. Statistical methods in psychology journals: Guidelines and explanations. *Am Psychologist* 1999; 54: 594-604.

APPENDIX I

Assessment of Study Quality

All studies were evaluated according to criteria of external, internal and statistical validity. External and internal validity were indexed by two key concepts each (see below).

External Validity: External validity denotes the generalisability of study results to populations, times or situations other than those in the study. It was evaluated in relation to sampling and case definition. External validity related to sampling (EV-S) was considered:

++ High when results can be generalized to the population of sleep disordered persons and this is achieved by random sampling. Random sampling might be modified according to specific needs or pre-existing knowledge (e.g. quota sampling, enriched sampling);
+ Satisfactory when results can be generalized to any somewhat smaller predefined population of sleep disordered persons (e.g. diagnosed patients, patients in a specific sleep laboratory) and this is achieved by random sampling;
0 Undetermined for any other attempt to achieve representativeness of the studied sample that does not constitute random sampling but is not completely controlled by the experimenter either, e.g, selection of consecutively referred patients;
- Unsatisfactory for studies lacking or not reporting any attempt to achieve representativeness.

Different sleep disorders require different diagnostic procedures, which often, but not always, include an overnight polysomnography and/or additional examinations. Case definitions or inclusion criteria that do not ensure diagnostic certainty to a reasonable extent will endanger generalisability to sleep disordered populations since the inclusion of false-positive and the exclusion of false-negative cases cannot be ruled out. Since standard guidelines or recommendations exist for the diagnosis of most sleep disorders [62], external validity related to case definition (EV-CD) was considered:

++ High when diagnostic procedures ensure the presence of the sleep disorder in the studied patients and the absence of the sleep disorder in the control group according to standard guidelines;
+ Satisfactory when at least the presence of the sleep disorder is diagnosed by standard procedures, and some attempt is undertaken to reasonably ensure the absence of the sleep disorder in the control group;
0 Undetermined when no guidelines or recommendations exist, or if diagnostic procedure employs some, but not all, of the diagnostic requirements;
- Unsatisfactory when minimal diagnostic requirements are not met.

Internal Validity: The internal validity of a study is the extent to which its design and conduct are likely to prevent systematic errors or biases. Internal validity was evaluated in relation to systematic differences in comparison groups other than the presence or absence of the sleep disorder (selection bias–IV-S), and in situational variables or care provided (performance bias–IV-P). Internal validity with regard to selection bias (IV-S) was considered:

++ High when patients and controls are drawn at random from a common predefined population;
+ Satisfactory when patients and controls are drawn at random from two different but reasonably similar populations, with (in smaller samples) or without (in larger samples) matching on relevant parameters;
0 Undetermined for any other attempt to minimize selection bias unless proven unsatisfactory;
- Unsatisfactory for any obvious differences between comparison groups that can be expected to correlate with the target measures and are not controlled statistically or otherwise.

Performance biases are systematic differences in care provided or procedures undertaken between patients and controls. Internal validity with regard to performance bias (IV-P) was considered:

++ High when all procedures (diagnostic or otherwise) are completely standardized and research personnel undertaking neuropsychological evaluations, scoring of sleep records or providing any other care are blind to the status (patient vs. control) of individual subjects;

+ Satisfactory when neuropsychological investigation as the target parameter is completely standardized and objectivity of conduct and scoring can be reasonably assumed;

0 Undetermined for any other attempt to minimize performance bias unless proven unsatisfactory;

- Unsatisfactory for any obvious differences in procedures that are likely to influence target measures.

Statistical Validity is the adequacy of the analysis as well as the correspondence between statistical results and scientific conclusions. Statistical validity is the only criterion in study assessment that is–apart from power considerations–not a fixed property of the study design and may thus be retrospectively changed by secondary analysis. Furthermore, statistical validity does not influence subsequent meta-analyses when only means and standard deviations are extracted from individual studies. It does however influence the interpretation of a single result from a single study. Owing to the possibility of retrospective correction, statistical validity will even in extreme cases not be judged as unsatisfactory but only as undetermined. Statistical validity was considered:

++ High when design and measures of the study correspond to the assumption of the test and type I errors are correct and type II errors are reasonably low;

+ Satisfactory when the above applies with the exception of type II errors;

0 Undetermined for any other studies.

APPENDIX II

Statistical Integration of Outcome Measures

Means and standard deviations were extracted for each outcome measure. Group differences were standardized using Hedges' adjusted effect size estimator $g_i=[(M_{1i}-M_{2i})/S_i][1-3/(4N_i-9)]$, where M_{1i} and M_{2i} denote the group means, S_i the pooled standard deviation, and N_i the pooled number of participants of both groups [79].

The standard error of g_i is $SE(g_i)=[N_i/N_1N_2+g_i^2/2(N_i-3.94)]^{1/2}$ and N_1 and N_2 the individual group sizes. Ninety-five percent confidence intervals for individual effect sizes were computed as $g\pm t_{crit}SE(g)$, where t_{crit} is the critical value of a t distribution with $N_1 + N_2 - 2$ degrees of freedom and a two-tailed α-level of 0.05.

Between-study heterogeneity was assessed with the heterogeneity statistic $Q=\Sigma vi(g_i-\Phi_{IV})2$, where $v_i=1/SE^2(g_i)$ and $\Phi_{IV}=(\Sigma v_i g_i)/(\Sigma v_i)$.

Q follows the distribution on $k-1$ degrees of freedom, where k is the number of studies contributing to the meta-analysis. Both the inverse variance fixed effects model and the DerSimonian and Laird's (1986) [80] random effects model were computed. For the fixed effects model individual effect sizes g_i were aggregated into a pooled effect size $\Phi_{IV}=(\Sigma v_i g_i)/(\Sigma v_i)$ with $SE(\Phi_{IV})=1/(\Sigma v_i)^{1/2}$, where $v_i=1/SE^2(g_i)$.

For the random effects model, individual effect sizes g_i were aggregated into a pooled effect size $\Phi_{DS}=(\Sigma w_i g_i)/(\Sigma w_i)$, with $SE(\Phi_{DS})=1/(\Sigma v_i)^{1/2}$, where $w_i=1/[SE^2(g_i)+\tau^2]$, and $\tau^2=\max\{[Q-(k-1)]/[\Sigma v_i-(\Sigma v_i^2)/\Sigma vi], 0\}$ with $Q=\Sigma v_i(g_i-\Phi_{IV})^2$ and $v_i=1/SE^2(g_i)$.

In case of homogeneity with $Q < df$, the fixed effects model and the random effects model will yield identical effect size estimates and standard errors. In the case of $Q > df$, the random effects model will give a broader confidence interval but assigns less weight to sample size than in the fixed effects model. Due to the difference in weighting, the random effects model sometimes yields less conservative estimates in situations, where large sample studies with relatively minor effect sizes are included as well as small sample studies with relatively large effect sizes (cf., Pool and Greenland, 1999).

APPENDIX III

Table 1
Perception and Attention

Study	Outcome Measure	Finding
Perception		
Bédard et al., 1991(a)	Hooper visual organization test	No difference vs. controls
Bédard et al., 1991(b)	Hooper visual organization test	No difference vs. controls
Bédard et al., 1991(a)	Visual matching test (Thurstone)	No difference vs. controls
Bédard et al., 1991(b)	Visual matching test (Thurstone)	No difference vs. controls
Knight et al., 1987	Graphesthesia	No difference vs. controls
Dani et al., 1996	Facial recognition (Benton)	Reduced performance vs. controls
Lee et al., 1999	Sensory motor task: correct	Improved performance vs. controls
Alertness, Simple Reaction Time (RT)		
Bonanni et al., 1999	Visual RT	Reduced performance vs. data base
Lee et al., 1999	Choice reaction time: 2 min	No difference vs. controls
Kotterba et al., 1998	RT	Reduced performance vs. controls
		Impaired performance vs. norms (mean percentile rank <30)
Weeß, 1996	CFF: threshold	No difference vs. controls
Rohmfeld et al., 1994(a)	CFF	No difference vs. controls
Rohmfeld et al., 1994(b)	CFF	No difference vs. controls
Verstraeten et al., 2000	Phasic alertness task	Two of 17 patients show impaired performance (percentile rank <5)
	Tonic alertness task	None of 17 patients show impaired performance (percentile rank <5)
Attentional Span, Short-term Attention		
Lee et al., 1999	Digit span forward (WAIS-R)	No difference vs. controls
Naëgelé et al., 1995	Digit forward	Reduced performance vs. controls
Verstraeten et al., 2000	Digit span forward	Five of 17 patients show impaired performance (percentile rank <5)
Knight et al., 1987	Digit span reversed (WAIS-R)	No difference vs. controls
Lee et al., 1999	Digit span reversed (WAIS-R)	No difference vs. controls
Naëgelé et al., 1995	Digit span reversed	Reduced performance vs. controls
Redline et al., 1997	Digit span reversed (WAIS-R)	Reduced performance vs. controls
Verstraeten et al., 2000	Digit span reversed	Four of 17 patients show impaired performance (percentile rank <5)
Knight et al., 1987	Digit span (WAIS-R)	No difference vs. controls
Lauer et al., 1998	Digit span	No difference vs. controls
Pietrini et al., 1998	Digit span	Reduced performance vs. controls
Greenberg et al., 1987	Digit span (WAIS-R)	Reduced performance vs. controls
Dani et al., 1996	Digit span	Reduced performance vs. controls
Borak et al., 1996	Digit span (WAIS)	Impaired performance vs. norms
Pietrini et al., 1998	Hiskey-Nebraska blocks	Reduced performance vs. controls
Lauer et al., 1998	Corsi block tapping task	No difference vs. controls
Naëgelé et al., 1995	Corsi block-tapping task	Reduced performance vs. controls
Naëgelé et al., 1995	Double encoding task: visual span	Reduced performance vs. controls
	Double encoding task: verbal span	Reduced performance vs. controls
	Double encoding task: double span	Reduced performance vs. controls
Focused Attention: Trail Making Test (TMT)		
Naëgelé et al., 1995	TMT A	No difference vs. controls
Redline et al., 1997	TMT A	No difference vs. controls
Phillips et al., 1994	TMT	No difference at baseline and 3-year follow-up vs. controls.
Zozula et al., 1998a	TMT A	No difference vs. controls
Kuo et al., 2000	TMT A	No difference vs. controls
Lauer et al., 1998	TMT	No difference vs. controls
Lee et al., 1999	TMT A	No difference vs. controls
Kotterba et al., 1998	TMT	Reduced performance vs. controls
		Unimpaired performance vs. norms (percentile rank >30)
Naëgelé et al., 1999	TMT A	Reduced performance vs. controls
Cassel et al., 1989	TMT	Impaired performance vs. norms
Walsleben et al., 1989	TMT A	Impaired performance vs. norms
Kotterba et al., 1997	TMT	32 of 40 patients showed impairment performance vs. norms (percentile rank <25)
Verstraeten et al., 2000	TMT A	Two of 17 patients show impaired performance (percentile rank <5)
Bédard et al., 1991(a)	TMT B	No difference vs. controls
Naëgelé et al., 1995	TMT B	No difference vs. controls
Redline et al., 1997	TMT B	No difference vs. controls
Greenberg et al., 1987	TMT B	No difference vs. controls
Kim et al., 1997	TMT B	No difference vs. controls
Kuo et al., 2000	TMT B	No difference vs. controls
Findley et al., 1991(a)	TMT B	No difference vs. controls

Table 1 (cont.)

Study	Outcome Measure	Finding
Focused Attention: Trail Making Test (TMT) (cont.)		
Lee et al., 1999	TMT B	No difference vs. controls
Findley et al., 1986(a)	TMT B	Unimpaired performance vs. norms
Bédard et al., 1991(b)	TMT B	Reduced performance vs. controls
Zozula et al., 1998a	TMT B	Reduced performance vs. controls
Naëgelé et al., 1999	TMT B	Reduced performance vs. controls
Findley et al., 1986(b)	TMT B	Impaired performance vs. norms
Roehrs et al., 1995	TMT B	Impaired performance vs. norms
Walsleben et al., 1989	TMT B	Impaired performance vs. norms
Verstraeten et al., 2000	TMT B	Three of 17 patients show impaired performance (percentile rank <5)
Focussed Attention: Symbol Digit Substitution Test (SDST)/Digit Symbol Substitution Test (DSST)		
Redline et al., 1997	SDST (Gilmore-Royer)	No differences vs. controls
Kim et al., 1997	SDST	No difference vs. controls
Stone et al., 1994	SDST	No difference vs. insomniac controls
		Unimpaired performance vs. norms
Walsleben et al., 1989	SDST	Marginally reduced performance vs. norms
Verstraeten et al., 2000	SDST	Two of 17 patients show impaired performance (percentile rank <5)
Bédard et al., 1991(a)	DSST (WAIS-R)	No difference vs. controls
Redline et al., 1997	DSST (WAIS-R)	No difference vs. controls
Phillips et al., 1994	DSST (WAIS-R)	No difference at baseline and 3-year follow-up vs. controls
Zozula et al., 1998a	DSST (WAIS-R)	No difference vs. controls
Dinges et al., 1998	DSST	No difference vs. controls
Bédard et al., 1991(b)	DSST (WAIS-R)	Reduced performance vs. controls
Roehrs et al., 1995	DSST	Unimpaired performance vs. norms
Borak et al., 1996	DSST (WAIS)	Impaired performance vs. norms
Focussed Attention: Cancellation Tests		
Bédard et al., 1991(a)	Letter cancellation	No difference vs. controls
Knight et al., 1987	Letter cancellation	No difference vs. controls
Naëgelé et al., 1995	Digit cancellation	No difference vs. controls
Redline et al., 1997	Letter cancellation: errors	No difference vs. controls
Kim et al., 1997	Digit cancellation	No difference vs. controls
Lauer et al., 1998	Letter cancellation	No difference vs. controls
Cassel et al., 1989	Digit cancellation	Unimpaired performance vs. norms
Bédard et al., 1991(b)	Letter cancellation	Reduced performance vs. controls
Greenberg et al., 1987	Letter cancellation	Reduced performance vs. controls
Borak et al., 1996	Bourdon-Wiersma test	Impaired performance vs. norms
Kotterba et al., 1997	Digit cancellation	23 of 40 patients show impaired performance vs. norms (percentile rank <25)
Divided Attention		
Rohmfeld et al., 1994(a)	Serial addition	No difference vs. controls
Rohmfeld et al., 1994(b)	Serial addition	No difference vs. controls
Weeß, 1996	Serial addition	No difference vs. controls
Redline et al., 1997	Digit subtraction: % correct responses	No difference vs. controls
Lee et al., 1999	Serial subtraction: correct responses	No difference vs. controls
Van Son et al., 2000	N-back continuous attention memory task	No difference vs. controls
Sloan et al., 1989	PASAT	Reduced performance vs. controls
Findley et al., 1991(a)	PASAT	Reduced performance vs. controls
Findley et al., 1986 (a)	PASAT: % correct responses	Unimpaired performance vs. norms
Findley et al., 1986(b)	PASAT: % correct responses	Impaired performance vs. norms
Weeß, 1996	Divided attention (Wiener test system)	No difference vs. controls
Kotterba et al., 1998	Divided attention (TAP, test battery for measuring attention)	No difference vs. controls
		Unimpaired performance vs. norms (mean percentile rank >30)
	Divided attention: % mistakes	Reduced performance vs. controls
Sloan et al., 1989	Visual tracking	Reduced performance vs. controls
	Auditory tracking	Reduced performance vs. controls
Selective Attention		
Bonanni et al., 1999	Visual complex reaction time	Reduced performance vs. data base
Camus et al., 1999	Dichotic listening task	Reduced performance vs. controls
Kotterba et al., 1998	Selective attention (Wiener Test System)	Reduced performance vs. controls
		Unimpaired performance vs. norms (mean percentile rank >30)
Rohmfeld et al., 1994(a)	Q11 (Wiener Test System): tasks solved	No difference vs. controls
Rohmfeld et al., 1994(b)	Q11 (Wiener Test System): tasks solved	Reduced performance vs. controls
Phillips et al., 1994	Stroop test	No difference at baseline and 3-year follow-up vs. controls

Table 1 (cont.)

Study	Outcome Measure	Finding
Selective Attention (cont.)		
Naëgelé et al., 1995	Stroop test: Interference	Reduced performance vs. controls
Kuo et al., 2000	Stroop test: Interference	Reduced performance vs. controls
Lauer et al., 1998	Stroop test	Reduced performance vs. controls
Verstraeten et al., 1996	Stroop test: A,B, C, and interference	No difference vs. insomniac controls
Verstraeten et al., 1997	Stroop test: A, B, C	Reduced performance vs. snoring controls
Sustained Attention		
Findley et al., 1986(a)	FCRRT	Unimpaired performance vs. norms
Findley et al., 1986(b)	FCRRT	Unimpaired performance vs. norms
Bédard et al., 1991(a)	FCRTT: RT, 5 repeated trials, 10 min	No difference vs. controls
	FCRTT: gaps, 5 repeated trials, 10 min	No difference vs. controls
	FCRTT: errors, 5 repeated trials, 10 min	No difference vs. controls
Bédard et al., 1991(b)	FCRTT: RT, 5 repeated trials, 10 min	Reduced performance vs. controls
	FCRTT: gaps, 5 repeated trials, 10 min	Reduced performance vs. controls
	FCRTT: errors, 5 repeated trials, 10 min	No difference vs. controls
Zozula et al., 1998b	FCRRT: 5 trials, gaps	Reduced performance vs. controls
	FCRRT: 5 trials, RT	Reduced performance vs. controls
Morrison et al., 1997	FCRRT: RT	Reduced performance vs. controls
	FCRRT: errors	Reduced performance vs. controls
	FCRRT: gaps	Reduced performance vs. controls
Findley et al., 1991(a)	FCRRT: RT	No difference vs. controls
Verstraeten et al., 1996	FCRTT: RT	Improved performance vs. insomniac controls
	FCRTT: errors	Reduced performance vs. insomniac controls
Verstraeten et al., 1997	FCRTT: RT	No difference vs. snoring controls
Muñoz et al., 2000	PVT: RT, 10 min (80-85 signals)	Reduced performance vs. controls
Barbé et al., 1998	PVT: RT, 10 min (80-85 signals)	Reduced performance vs. controls
	PVT: reaction fatigue, 10 min (80-85 signals)	No difference vs. controls
Chugh et al., 1998	PVT: lapses and transformed lapses	Reduced performance vs. snoring controls
	PVT: fastest 10% reactions	No difference vs. snoring controls
	PVT: slowest 10% reactions	Reduced performance vs. snoring controls
Dinges et al., 1998	PVT: transformed lapses	Reduced performance vs. controls
Redline et al., 1997	CPT: d', 10 min, first 2 min	No difference vs. controls
	CPT: d', 10 min, last 2 min	Reduced performance vs. controls
Stone et al., 1994	CPT: RT	No difference vs. insomniac controls Unimpaired performance vs. norms
	CPT: commission errors	No difference vs. insomniac controls Unimpaired performance vs. norms
	Stone et al., 1994 (cont.)	CPT: omission errors No difference vs. insomniac controls Unimpaired performance vs. norms
Randerath et al., 2000	Sustained attention test: gaps	Reduced performance vs. controls
Kotterba et al., 1998	Continuous attention	Reduced performance vs. controls Impaired performance vs. norms (mean percentile rank <30)
Rohmfeld et al., 1994(a)	Modified sustained attention test	No difference vs. controls
Rohmfeld et al., 1994(b)	Modified sustained attention test	No difference vs. controls
Vigilance		
Weeß, 1996	Müggeburg test (Mackworth clock): correct, 66 min	Reduced performance vs. controls
	Müggeburg test (Mackworth clock): % missing, 66 min	Reduced performance vs. controls
Rohmfeld et al., 1994(a)	Müggeburg test (Mackworth clock): 66 min	No difference vs. controls
Rohmfeld et al., 1994(b)	Müggeburg test (Mackworth clock): 66 min	No difference vs. controls
Sauter et al., 2000	Quatember Maly (Mackworth clock): correct, 2 x 30 min	Mean score similar to mean of a normal population, (1/3 of patients with percentile rank <25)
Sauter et al., 2000(b)	Quatember Maly (Mackworth clock): correct 2 x 30 min	Mean score similar to mean of a normal population, (1/3 of patients with percentile rank <25)
Van Son et al., 2000	Parasuraman vigilance task	No difference vs. controls
Kotterba et al., 1998	Vigilance test	No difference vs. controls Unimpaired performance vs. norms (percentile rank >30)
	Vigilance test: % mistakes	Reduced performance vs. controls
Kotterba et al., 1997	Vigilance test	Seven of 40 patients show impaired performance vs. norms (percentile rank <25%)
Schulz et al., 1997	CFF (3 hours), modified: threshold	Reduced treshold vs. treated SRBD patients
	CFF (3 hours), modified: coefficient of variation	No difference vs. treated SRBD patients

Table 2
Motor Functions

Study	Outcome Measure	Finding
Psychomotor Speed		
Knight et al., 1987	Finger tapping: right and left hand	No difference vs. controls
Phillips et al., 1994	Finger Tapping	No difference at baseline and 3-year follow-up vs. controls.
Sloan et al., 1989	Finger tapping	No difference vs. controls
Verstraeten et al., 1996	Finger tapping: dominant hand	No difference vs. insomniac controls
	Finger tapping: non-dominant hand	No difference vs. insomniac controls
Verstraeten et al., 1997	Finger tapping: dominant hand	No difference vs. snoring controls
	Finger tapping: non-dominant hand	No difference vs. snoring controls
Roehrs et al., 1995	Finger tapping	No difference vs. norms
Verstraeten et al., 2000	Finger tapping: dominant hand	None of 17 patients show impaired performance (percentile rank <5)
Stone et al., 1994	Digit copying task	No difference vs. insomniac controls
		Unimpaired performance vs. norms
Lee et al., 1999	Sensory motor task: reaction time	Reduced performance vs. controls
Manual Dexterity and Speed		
Kim et al., 1997	Grooved pegboard	No difference vs. controls
Kuo et al., 2000	Grooved pegboard: dominant hand	No difference vs. controls
	Grooved pegboard: non-dominant hand	No difference vs. controls
Bédard et al., 1991(a)	Purdue pegboard: dominant hand	Reduced performance vs. controls
Bédard et al., 1991(b)	Purdue pegboard: dominant hand	Reduced performance vs. controls.
Greenberg et al., 1987	Purdue pegboard: both hands	Reduced performance vs. controls
	Purdue pegboard: left hand	Reduced performance vs. controls
	Purdue pegboard: right hand	Reduced performance vs. controls
Zozula et al., 1998a	Purdue pegboard: both hands	Reduced performance vs. controls
Verstraeten et al., 1996	Purdue pegboard: dominant hand	No difference vs. insomniac controls
	Purdue pegboard: non dominant hand	No difference vs. insomniac controls
Verstraeten et al., 1997	Purdue pegboard: dominant hand	Reduced performance vs. snoring controls
	Purdue pegboard: non dominant hand	No difference vs. snoring controls
Walsleben et al., 1989	Purdue pegboard	No difference vs. norms
Verstraeten et al., 2000	Purdue pegboard: non dominant hand	Three of 17 patients show impaired performance (percentile rank <5)

Table 3
Driving Simulation

Study	Outcome Measure	Finding
Ingram et al., 1994	Steer Clear, 30 min	No difference vs. controls
Findley et al., 1999	Steer Clear, 30 min: transformed total errors	No difference vs. controls
Findley et al., 1995(a)	Steer Clear, 30 min: % hits	Reduced performance vs. controls
Findley et al., 1995(b)	Steer Clear, 30 min: % hits	Reduced performance vs. controls
Muñoz et al., 2000	Steer Clear: 30 min, % hits	Reduced performance vs. controls
Barbé et al., 1998	Steer Clear, 30 min: % hits	Reduced performance vs. controls
Findley et al., 1991(b)	Steer Clear, 30 min: % hits	Reduced performance vs. controls
George et al., 1996	Divided attention driving test, 20 min:	
	No. of out of bounds	Reduced performance vs. controls
	No. of correct responses	Reduced performance vs. controls
	Response time	Reduced performance vs. controls
	Tracking errors	Reduced performance vs. controls
Dinges et al., 1998	Divided attention driving task: intercept	Reduced performance vs. controls
Risser et al., 2000	Driving simulator, 60 min: lane position variability	Reduced performance vs. controls
	Speed variability	Reduced performance vs. controls
	Steering rate variability	Reduced performance vs. controls
	Frequency of crashes	Reduced performance vs. controls
Findley et al., 1989(a)	Computer simulator, 30 min: No. hits	Reduced performance vs. controls
Findley et al., 1989(b)	Film driving simulator, 22 min % correct responses (highway driving)	Reduced performance vs. controls
	Film driving simulator, 22 min % correct responses (city/rural driving)	Reduced performance vs. controls
Juniper et al., 2000	Driving simulator: 30 min, all, near or far road visible, reaction time	Reduced performance vs. controls
	Off road events	Reduced performance vs. controls
Büttner et al., 2000	Carsim driving simulator: 30 min, No. of tracking errors	Reduced performance vs. controls
	No. of deviations from an ideal track	Reduced performance vs. controls

Table 4
Constructional Performance

Study	Outcome Measure	Finding
Copying		
Bédard et al., 1991(a)	Rey-Osterreith Figure	No difference vs. controls
Bédard et al., 1991(b)	Rey-Osterreith Figure	Reduced performance vs. controls
Greenberg et al., 1987	Bender Gestalt Test	Reduced performance vs. controls
Kales et al., 1985	Bender Gestalt Test: Z-Score	Of 50 patients, 12 (24%) showed mild to severe impairment (Z>72), 26 (52%) showed suspected impairment (72>Z>50)
Drawing		
Pietrini et al., 1998	Extended range drawing test	Reduced performance vs. controls
Building and Assembling		
Bédard et al., 1991(a)	WAIS-R: Block design	No difference vs. controls
Knight et al., 1987	WAIS-R: Block design	No difference vs. controls
Greenberg et al., 1987	WAIS-R: Block design	No difference vs. controls
Zozula et al., 1998a	WAIS-R: Block design	No difference vs. controls
Walsleben et al., 1989	WAIS-R: Block design	Unimpaired performance vs. norms
Borak et al., 1996	WAIS: Block design	Impaired performance vs. norms
Bédard et al., 1991(b)	WAIS-R: Block design	Reduced performance vs. controls
Bédard et al., 1991(a)	WAIS-R: Object assembly	No difference vs. controls
Bédard et al., 1991(b)	WAIS-R: Object assembly	Reduced performance vs. controls

Table 5
Memory, Learning and Forgetting

Study	Outcome Measure	Finding
Immediate Recall–Visual		
Knight et al., 1987	WMS: figural memory	No difference vs. controls
Greenberg et al., 1987	WMS: figural memory	No difference vs. controls
Findley et al., 1991(a)	WMS: figural memory	No difference vs. controls
Findley et al., 1986(a)	WMS: figural memory	Unimpaired performance vs. norms
Findley et al., 1986(b)	WMS: figural memory	Unimpaired performance vs. norms
Pietrini et al., 1998	WMS: figural memory	Reduced performance vs. controls
Klonoff et al., 1987	BVRT: errors	No difference vs. controls scheduled for surgery
Verstraeten et al., 1996	BVRT: correct	No difference vs. insomniac controls
	BVRT: errors	No difference vs. insomniac controls
Verstraeten et al., 1997	BVRT: correct	No difference vs. snoring controls
	BVRT: errors	No difference vs. snoring controls
Borak et al., 1996	BVRT	Nine of twenty patients scored below norms
Bédard et al., 1991(a)	Rey-Osterrieth figure	Reduced performance vs. controls
Bédard et al., 1991(b)	Rey-Osterrieth figure	Reduced performance vs. controls
Immediate Recall–Verbal and Story Recall		
Lee et al., 1999	CVLT: recall of Trial 1	No difference vs. controls
Bédard et al., 1991(a)	WMS: logical memory	No difference vs. controls
Bédard et al., 1991(b)	WMS: logical memory	No difference vs. controls
Knight et al., 1987	WMS: logical memory	No difference vs. controls
Greenberg et al., 1987	WMS: logical memory	No difference vs. controls
Findley et al., 1991(a)	WMS: logical memory	No difference vs. controls
Findley et al., 1986(a)	WMS: logical memory	Unimpaired performance vs. norms
Findley et al., 1986(b)	WMS: logical memory	Unimpaired performance vs. norms
Borak et al., 1996	AVLT: recall of Trial 1	Performance “highly abnormal” vs. norms
Learning–Verbal and Visual		
Redline et al., 1997	CVLT: list learning	No difference vs. controls
Knight et al., 1987	WMS: Associate learning	No difference vs. controls
Kim et al., 1997	AVLT: learning	No difference vs. controls
Borak et al., 1996	AVLT: learning	Performance “highly abnormal” vs. norms
Naëgelé et al., 1995	Verbal learning test: selective reminding	Reduced performance vs. controls
	Visual learning test	Reduced performance vs. controls
Retrieval: Delayed Recall – Short-term Retention with Interference		
Knight et al., 1987	WMS: associate learning, 30 sec	No difference vs. controls
	WMS: logical memory, 30 sec	No difference vs. controls
	WMS: figural memory, 30 sec	No difference vs. controls
Stone et al., 1994	Unclustered word list	No difference vs. insomniac controls
		Unimpaired performance vs. norms
	Clustered word list	No difference vs. insomniac controls
		Unimpaired performance vs. norms

Table 5 (cont.)

Study	Outcome Measure	Finding
Free Recall–Long Delay: Visual		
Greenberg et al., 1987	WMS: figural memory	No difference vs. controls
Findley et al., 1991(a)	WMS: figural memory, 30 min delay	No difference vs. controls
Findley et al., 1986(a)	WMS: figural memory, 30 min delay	Unimpaired performance vs. norms
Berry et al., 1990	WMS: figural memory	Reduced performance vs. controls
Pietrini et al., 1998	WMS: figural memory	Reduced performance vs. controls
Findley et al., 1986(b)	WMS: figural memory, 30 min delay	Impaired performance vs. norms
Bédard et al., 1991(a)	Rey-Osterrieth figure: 60 min delay	No difference vs. controls
Phillips et al., 1994	Rey-Osterrieth figure	No difference at baseline, at 3-year follow-up or in time course vs. controls
Bédard et al., 1991(b)	Rey-Osterrieth figure: 60 min delay	Reduced performance vs. controls
Zozula et al., 1998a	Rey-Osterrieth figure	Reduced performance vs. control
Free Recall–Long Delay: Verbal		
Redline et al., 1997	CVLT	No difference vs. controls
Bédard et al., 1991(a)	WMS: logical memory, 60 min delay	No difference vs. controls
Berry et al., 1990	WMS: logical memory	No difference vs. controls
Greenberg et al., 1987	WMS: logical memory	No difference vs. controls
Findley et al., 1991(a)	WMS: logical memory, 30 min delay	No difference vs. controls
Findley et al., 1986(a)	WMS: logical memory, 30 min delay	Unimpaired performance vs. norms
Bédard et al., 1991(b)	WMS: logical memory, 60 min delay	Reduced performance vs. controls.
Findley et al., 1986(b)	WMS: logical memory, 30 min delay	Impaired performance vs. norms
Forgetting Visual		
Naëgelé et al., 1995	Visual learning test: % forgetting, 30 min delay	No difference vs. controls
Greenberg et al., 1987	WMS: figural memory, % retained	No difference vs. controls
Forgetting Verbal		
Naëgelé et al., 1995	Verbal learning test: % forgetting, 30 min delay	No difference vs. controls
Greenberg et al., 1987	WMS: logical memory, % retained	No difference vs. controls
Kim et al., 1997	AVLT: % retained, 30 min delay	No difference vs. controls
Recognition		
Kim et al., 1997	AVLT: recognition	No difference vs. controls
Complex/Multiple Measures		
Knight et al., 1987	CVLT	No difference vs. controls
Kim et al., 1997	Factor Memory	No difference vs. controls
Kotterba et al., 1998	Verbal learning and memory test	No difference vs. controls
Phillips et al., 1994	WMS: logical memory	No difference at baseline, 3-year follow-up or in time course vs. controls.
	WMS: figural memory	No difference at baseline, 3-year follow-up or in time course vs. controls.
Roehrs et al., 1995	WMS: figural memory	Unimpaired performance vs. norms
	WMS: logical memory	Impaired performance vs. norms
Lauer et al., 1998	Modified CVLT	No difference vs. controls
Walsleben et al., 1989	CVLT	Unimpaired performance vs. norms
	WMS	Unimpaired performance vs. norms
Lojander et al., 1999(a)	WMS	Unimpaired performance vs. norms
Lojander et al., 1999(b)	WMS	Unimpaired performance vs. norms
Lojander et al., 1999(c)	WMS	Unimpaired performance vs. norms
Lojander et al., 1999(d)	WMS	Unimpaired performance vs. norms
Dani et al., 1996	WMS	Reduced performance vs. controls
Kales et al., 1985	Memory impairment (WMS <86)	Of 22 patients, 11 (52%) showed impaired memory
Sloan et al., 1989	Delayed memory	Reduced performance vs. controls
Naëgelé et al., 1999	Visual short-term memory	Reduced performance vs. controls
Other Memory Tasks		
Stone et al., 1994	Symbol-digit substitution recall=Incidental learning	No difference vs. insomniac controls after Bonferroni correction and norms
	Category shifts /unclustered words=spontaneous use of effortful encoding strategies	No difference vs. insomniac controls after Bonferroni correction and norms
Lee et al., 1999	Spatial working memory task: correct	No difference vs. controls
	Spatial working memory task: reaction time	No difference vs. controls
Redline et al., 1997	Pursuit rotor learning (3 trials): % change in time	No differences vs. controls
Kuo et al., 2000	Working memory index (WAIS-III)	No difference vs. controls
Dinges et al., 1998	Probed recall memory test	No difference vs. controls
Naëgelé et al., 1999	Visual constructive procedural learning	Reduced performance vs. controls
	Working memory	Reduced performance on 3 out of 6 tests vs. controls
Kales et al., 1985	Long-term memory impairment (Information (WAIS) or personal and current information (WMS) <1 SD of norms)	Of 50 patients, 16 (32%) showed an impaired long-term memory

Table 6
Concept Formation, Reasoning, and Executive Functions

Study	Outcome Measure	Finding
Concept Formation in Visual Format		
Verstraeten et al., 1996	Raven progressive matrices: short form	No difference vs. insomniac controls
Verstraeten et al., 1997	Raven progressive matrices: short form	No difference vs. snoring controls
Klonoff et al., 1987	Halstead category test: short form, errors	No difference vs. controls scheduled for surgery
Lauer et al., 1998	Raven progressive matrices	Reduced performance vs. controls
Roehrs et al., 1995	Category Test	Impaired performance vs. norms
Concept Formation in Verbal Format		
Phillips et al., 1994	Similarities (WAIS-R)	No difference at baseline, 3-year follow-up or in time course vs. controls
Bédard et al., 1991(a)	Similarities (WAIS-R)	No difference vs. controls
Bédard et al., 1991(b)	Similarities (WAIS-R)	No difference vs. controls.
Sort and Shift		
Lee et al., 1999	WCST: perseverative errors	Reduced performance vs. controls
	WCST: categories	No difference vs. controls
Naëgelé et al., 1995	Modified WCST: % perseverative errors	Reduced performance vs. controls
	Modified WCST: categories	No difference vs. controls
	Modified WCST: errors	No difference vs. controls
Redline et al., 1997	WCST: perseverative errors	Reduced performance vs. controls
Naëgelé et al., 1999	WCST	Reduced performance vs. controls
Reasoning		
Bédard et al., 1991(a)	Comprehension (WAIS-R)	No difference vs. controls
Bédard et al., 1991(b)	Comprehension (WAIS-R)	No difference vs. controls
Naëgelé et al., 1995	20-question procedure	No difference vs. controls
Stone et al., 1994	Generating an optimal telegram: Telegram 1	No difference vs. insomniac controls
		Unimpaired performance vs. norms
	Telegram 2	No difference vs. insomniac controls
		Unimpaired performance vs. norms
Borak et al., 1996	Arithmetic (WAIS-R)	Reduced performance vs. norms
Bédard et al., 1991(a)	Picture arrangement (WAIS-R)	Reduced performance vs. controls
Bédard et al., 1991(b)	Picture arrangement (WAIS-R)	Reduced performance vs. controls
Borak et al., 1996	Picture arrangement (WAIS-R)	Reduced performance vs. norms
	Picture completion (WAIS-R)	Unimpaired performance vs. norms
Executive Functions		
Stone et al., 1994	Porteus Maze Test	No difference in gender-adjusted scores vs. insomniac controls
		Unimpaired performance vs. norms
Knight et al., 1987	Maze tracing speed	No difference vs. controls
Bédard et al., 1991(a)	Mazes	Reduced performance vs. controls
Bédard et al., 1991(b)	Mazes	Reduced performance vs. controls.
Lee et al., 1999	Tower puzzles: number solved	No difference vs. controls
	Tower puzzles: errors	No difference vs. controls
Naëgelé et al., 1995	Tower of Toronto: 3 disks	Reduced performance vs. controls
	Tower of Toronto: 4 disks	No difference vs. controls

Table 7
Verbal Functions and Language Skills

Study	Outcome Measure	Finding
Verbal Fluency		
Knight et al., 1987	COWAT	No difference vs. controls
Greenberg et al., 1987	COWAT	No difference vs. controls
Kim et al., 1997	COWAT	No difference vs. controls
Lee et al., 1999	COWAT	No difference vs. controls
Naëgelé et al., 1995	Verbal fluency	No difference vs. controls
Bédard et al., 1991(a)	Verbal fluency	No difference vs. controls
Bédard et al., 1991(b)	Verbal fluency	Reduced performance vs. controls
Walsleben et al., 1989	Verbal fluency	No differnce vs. norms
Verbal Expression–Vocabulary		
Knight et al., 1987	WAIS-R: Vocabulary	No difference vs. controls
Greenberg et al., 1987	WAIS-R: Vocabulary	No difference vs. controls

Table 7 (cont.)

Study	Outcome Measure	Finding
Verbal Academic Skills–Knowledge Aquisition and Retention		
Greenberg et al., 1987	WAIS-R: Information	No difference vs. controls
Ingram et al., 1994	WAIS-R: Information	No difference vs. controls
Walsleben et al., 1989	WAIS-R: Information	No difference vs. norms
Borak et al., 1996	WAIS: Information	No difference vs. norms
Verbal Expression–Confrontation Naming		
Knight et al., 1987	Boston Naming Test	No difference vs. control

Table 8
Composite Measures

Study	Outcome Measure	Finding
Full Scale IQ		
Bédard et al., 1991(a)	WAIS-R	No difference vs. controls
Klonoff et al., 1987	WAIS-R	No difference vs. controls scheduled for surgery
Borak et al., 1996	WAIS	Unimpaired performance vs. norms
Bédard et al., 1991(b)	WAIS-R	Reduced performance vs. controls
Zozula et al., 1998a	WAIS-R	Reduced performance vs. controls
Pietrini et al., 1998	WAIS	Reduced performance vs. controls
Verbal IQ		
Bédard et al., 1991(a)	WAIS-R: verbal IQ	No difference vs. controls
Bédard et al., 1991(b)	WAIS-R: verbal IQ	No difference vs. controls
Berry et al., 1990	WAIS-R: verbal IQ	No difference vs. controls
Berry et al., 1987	WAIS-R: verbal IQ	No difference vs. controls
Lojander et al., 1999(a)	WAIS: verbal IQ	Unimpaired performance vs. norms
Lojander et al., 1999(b)	WAIS: verbal IQ	Unimpaired performance vs. norms
Lojander et al., 1999(c)	WAIS: verbal IQ	Unimpaired performance vs. norms
Lojander et al., 1999(d)	WAIS: verbal IQ	Unimpaired performance vs. norms
Pietrini et al., 1998	WAIS: verbal IQ	Reduced performance vs. controls
Performance IQ		
Bédard et al., 1991(a)	WAIS-R: performance IQ	No difference vs. controls
Berry et al., 1987	WAIS-R: performance IQ	No difference vs. controls
Lojander et al., 1999(a)	WAIS: performance IQ	Unimpaired performance vs. norms
Lojander et al., 1999(b)	WAIS: performance IQ	Unimpaired performance vs. norms
Lojander et al., 1999(c)	WAIS: performance IQ	Unimpaired performance vs. norms
Lojander et al., 1999(d)	WAIS: performance IQ	Unimpaired performance vs. norms
Bédard et al., 1991(b)	WAIS-R: performance IQ	Reduced performance vs. controls
Berry et al., 1990	WAIS-R: performance IQ	Reduced performance vs. controls
Pietrini et al., 1998	WAIS: performance IQ	Reduced performance vs. controls
Other		
Findley et al., 1991(a)	WAIS-R: Vocabulary and Block design	No difference vs. controls
Findley et al., 1986 (a)	WAIS-R: Vocabulary and Block design	Unimpaired performance vs. norms
Findley et al., 1986 (b)	WAIS-R: Vocabulary and Block design	Unimpaired performance vs. norms
Phillips et al., 1994	MMSE	No difference at baseline, 3-year follow-up or in time course vs. controls
Berry et al., 1990	MMSE	No difference vs. controls
Kuo et al., 2000	Processing speed index (WAIS-III)	No difference vs. controls
Kim et al., 1997	Factor Psychomotor efficiency	Reduced performance vs. controls
Sloan et al., 1989	Shipley IQ estimate	Reduced performance vs. controls
Kales et al., 1985	Cognitive impairment (WAIS-Verbal IQ – WAIS-Performance IQ >15)	Of 28 patients, 11 (39%) showed impairment