in press, Journal of Clinical and Experimental Neuropsychology

A Meta-Analytic Review of Verbal Fluency Deficits in Depression

Julie. D. Henry and John. R. Crawford

Department of Psychology

King's College

University of Aberdeen

Abbreviated Title: Meta-analysis of fluency in depression

Word count (excluding references): 8264

Correspondence to: Julie D. Henry, Department of Psychology, King's College, University of Aberdeen, AB24 3HN (e-mail: j.d.henry@abdn.ac.uk; Telephone; (0) 1224 273483; Fax (0) 1224 273426).

ABSTRACT

A prominent view in the neuropsychological literature is that depression is particularly associated with deficits in executive control processes. A meta-analysis of 42 studies with 2206 participants was therefore conducted to investigate the sensitivity of tests of verbal fluency to the presence of this disorder, as there is a great deal of evidence that these measures are valid markers of executive dysfunction. When the methodology adopted by other meta-analytic reviews was employed, semantic fluency deficits were found to be substantially larger than phonemic fluency deficits. However, when a more rigorous method of meta-analysis was adopted, this indicated that the measures are in fact broadly equivalent in their sensitivity to depression, as has been found for patients with focal frontal lobe lesions. However, in contrast to patients with focal frontal lobe injuries, neither deficit qualified as a differential deficit relative to psychomotor speed. Therefore, for patients with depression, deficits on tests of phonemic and semantic fluency may not reflect executive dysfunction, but a more generalised impairment. Evidence is presented that tests of phonemic and semantic fluency may aid in the differential diagnosis of patients with depression and those in the early stages of dementia of the Alzheimer's type.

INTRODUCTION

It is now accepted that depression is associated with a number of neurocognitive deficits (Christensen, Griffiths, Mackinnon, & Jacomb, 1997), but there remains disagreement as to which aspects of cognition are particularly affected. This problem has been exacerbated by the fact that the disorder is often characterised by generalised intellectual impairment, which may reflect psychomotor retardation (a generalised slowing of mental processes), and/or an attentional deficit. Indeed, the level of generalised cognitive impairment associated with depression can be so pronounced that in the elderly the condition is misdiagnosed as dementia (Flint & Eastwood, 1988; Wertheimer, 1991). However, whilst patients often perform poorly on virtually all measures of cognition, it has been suggested that executive dysfunction may be one of the most prominent disturbances associated with the disorder (Fossati, Ergis, & Allilaire, 2002; Kaiser et al., 2003; Veiel, 1997).

Executive functioning is hypothesised to be responsible, not for basic cognitive processes, but for the set of behavioural competencies that integrate these capacities (Della Sala, Gray, Spinnler, & Trivelli, 1998), and thus permits contextually sensitive, flexible responses. Aspects of executive function include self-directed planning and strategy formation, future-orientated, goal-directed and non-habitual behaviour (Crawford & Henry, in press; Perret, 1974; Phillips, 1997; Stuss & Benson, 1986). Consistent with this perspective, patients with depression often perform poorly on tests designed to capture executive dysfunction, including the Wisconsin Card Sorting Test (WCST; Franke et al., 1993; Pendleton Jones, Henderson, & Welch, 1988) tests of phonemic and semantic fluency (Brown, Scott, Bench, & Dolan, 1994; Trichard et al., 1995) and the stroop interference test (Nathan, Wilkinson, Stammers, & Low, 2001). Moreover, neuropathologically there is evidence of fronto-thalamic striatal involvement in depression (see; Rogers, Bradshaw,

Pantellis, & Phillips, 1998). Since there is a great deal of evidence that executive processes rely heavily upon the intact functions of frontal structures (see, e.g.; Shallice, 1988; Stuss & Benson, 1986) the presence of frontal abnormalities would therefore suggest that deficits in this aspect of cognition should be especially marked.

However, fronto-executive hypotheses have been offered to account for the cognitive and behavioural disturbances seen in a vast array of neurological and psychiatric disorders, in addition to depression. If we are to avoid engaging in what David (1992) has termed "Frontal Lobology: Psychiatry's new pseudoscience", rigorous standards of proof must be applied when evaluating these hypotheses. Thus, given that generalised deficits are associated with depression, an important stage in evaluating an executive hypothesis is to test whether any observed deficits on executive tasks qualify as differential deficits (Crawford & Henry, in press; Laws, 1999; Miller, 1984). A deficit on an executive measure is not by itself sufficient to infer the presence of a differential executive deficit; instead, it must be shown that the executive deficit is in excess of the averaged performance deficit across a range of other cognitive tasks that are not considered to impose heavy executive demands.

Evidence for a Differential Deficit in Executive Functioning

Whilst a number of meta-analytic reviews have described how depression impacts upon cognitive functioning, they have yielded inconsistent results with regard to whether the disorder is particularly characterised by disturbances in executive control processes. Zakzanis et al. (1998) found that for patients with major depression, although deficits on several measures of executive function, such as semantic fluency, the Stroop interference, and phonemic fluency were above the median effect size, deficits upon the WCST were substantially below this value. Moreover, deficits

upon all executive measures were substantially smaller than those for measures of episodic, declarative memory. However, in another meta-analytic review, Veiel (1997) found that when 13 studies of major depression were pooled, putative measures of executive function were particularly disrupted. The largest deficit was associated with a category termed 'mental flexibility and control', which collapsed effects across measures of the Trail Making Test Part B and the interference condition of the Stroop (Glass's Δ = 2.00). However, phonemic fluency was associated with a substantially smaller mean effect (.55), which actually constituted the second smallest deficit of all ten of the cognitive categories assessed. Thus, the prominence of deficits on measures of executive functioning relative to measures presumed to make only minimal executive demands remains unclear.

It should be noted, however, that the generalisability of both these meta-analytic studies' results is limited as relatively few effects contributed to each respective mean (Veiel, 1997; Zakzanis et al., 1998). In Veiel's (1997) study only three effects contributed to each of the means for 'mental flexibility/control' and verbal fluency, whilst in Zakzanis et al.'s (1998) study, of 74 'mean' effects, over half were derived from two effects or less. A larger fixed-effects meta-analysis was conducted by Christensen et al. (1997), in which less restrictive eligibility criteria were employed. The cognitive deficits associated with primary depression in general were investigated, and thus, in addition to major depression, this study included patients with disorders such as dysthymia and bipolar depression. Whilst the largest deficits were associated with tests that required problem solving, memory and speed, virtually all measures were impaired, and were on average 0.63 standard deviations from controls. Relative to this 'general' level of impairment, the deficit for executive function ('goal formation and planning') was only slightly larger (.73). Christensen et al.

(1997) argued that the findings did not suggest that depression is particularly characterised by deficits in effortful processing, but that there may be a deficit in speed or attention.

However, it is problematic that both Christensen et al. (1997) and Veiel (1997) collapsed effects across different tests to create a composite 'executive' measure. This is because there are ambiguities with regard to what actually constitutes a valid measure of executive function and because of evidence of fractionation of executive processes (see; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Indeed, in Christensen et al.'s (1997) study, despite being one of the best validated measures of executive functioning, phonemic fluency was categorised as a test of verbal ability along with measures of aphasia, confrontation naming and the National Adult Reading Test (NART; Nelson, 1982). Another difficulty with this methodology is that even if valid measures of executive functioning are collapsed to create a composite measure, it remains unclear whether executive functioning in depression is compromised per se. The striatum and frontal cortex are linked by at least five parallel but independent circuits (Alexander, DeLong, & Strick, 1986). If fronto-striatal circuits are implicated in the etiology of depression, differential impairment in these prefrontal 'loops' would presumably cause deficits on some executive measures, but not others.

Indeed, some studies have found that different measures of executive functioning are not affected equivalently by depressive illness. Jones, Henderson and Welch (1988), for instance, reported that prior to electroconvulsive therapy (ECT), severely depressed patients tested on a diverse range of measures designed to capture executive dysfunction were very impaired on those that required cognitive shifting, such as the WCST and the 'Go-No Go' task, but performance on a test of phonemic fluency was preserved. Fossati et al. (1999) found that young inpatients with unipolar depression performed significantly worse than controls on semantic fluency, but not phonemic fluency, Cognitive Estimates, or the WCST. Degl'Innocenti, Agren and Backman (1998)

also reported that whilst phonemic fluency and the Stroop were significantly impaired, this was true for only four of the 11 dependent measures for the WCST. Moreover, the deficit on the Stroop was equivalent for all three conditions (i.e. the interference condition was not disproportionately impaired), consistent with there being a reduction in information processing speed, but not a breakdown in inhibitory processes. Degl'Innocenti et al. (1998) concluded that; "The overall pattern of results suggests that depression may affect various executive functions in a differentiated manner." (p.182).

Verbal Fluency Deficits in Depression

In an attempt to resolve whether depression is associated with a differential executive deficit, verbal fluency performance has been studied extensively. This is particularly important given that Norris et al. (1995) have suggested that performance on tests of verbal fluency may be especially sensitive to depressive illness as there are clear similarities between the cognitive demands of fluency measures, and the deficits typically associated with the disorder, including the capacity for sustained attention, concentration, retrieval and speed. Tests of verbal fluency are amongst the most widely employed measures used to assess cognitive functioning following neurological damage, and involve associative exploration and retrieval of words based on phonemic or semantic criteria (phonemic and semantic fluency, respectively), usually conducted in the setting of a time constraint. Thus, whilst for phonemic fluency participants are asked to generate as many words as possible beginning with a specified letter (e.g. F), for semantic fluency search is constrained by a specified category (e.g. animals). These measures are considered to impose comparable demands upon executive or supervisory processes because both require efficient organisation of verbal retrieval and recall, as well as self-monitoring aspects of cognition (the participant must keep track of

responses already given), effortful self-initiation, and inhibition of responses when appropriate (Crawford & Henry, in press; Ruff, Light, Parker, & Levin, 1997). However, whilst phonemic fluency requires the creation of search strategies based primarily on lexical representations, tests of semantic fluency require searching for semantic extensions of a target superordinate, and thus depend intrinsically upon the integrity of semantic associations within the lexicon (Rohrer, Salmon, Wixted, & Paulsen, 1999). Deficits on tests of semantic fluency may therefore reflect problems with semantic memory, and not executive dysfunction.

In a recent meta-analytic review that included 31 studies and 1791 participants, Henry and Crawford (in press-a) investigated the relative magnitude of cognitive deficits upon tests of phonemic and semantic fluency for patients with focal cortical lesions. The pattern of results suggested that whilst the two types of fluency impose comparable demands upon executive processes, semantic fluency is relatively more dependent upon the integrity of semantic memory. Henry and Crawford (in press-a) found that focal frontal lobe injuries were associated with equivalent phonemic and semantic fluency deficits ($\underline{r}$s = .52 and .54 respectively). Since frontal structures are particularly implicated in executive functioning (see; Shallice, 1988; Stuss & Benson, 1986), a pattern of comparable impairment upon tests of phonemic and semantic fluency for patients with depression may therefore reflect executive dysfunction if, as was found for frontal patients (but not for non-frontal patients), verbal fluency deficits qualify as differential deficits relative to verbal intelligence and psychomotor speed.

However, semantic fluency was more impaired following focal temporal damage ($\underline{r}$ = .61), and this deficit was substantially larger than the corresponding phonemic fluency deficit ($\underline{r}$ = .44). Since there is a great deal of evidence that temporal structures are the neural substrates particularly responsible for semantic memory (see; Martin & Chao, 2001), this was presumed to reflect the

greater reliance of semantic fluency upon the integrity of semantic memory. Comparison of the relative magnitude of deficits on phonemic and semantic fluency may therefore be used to draw inferences regarding the prominence of executive deficits and semantic store degradation, respectively.

However, as for other measures of executive function, empirical investigation has not consistently found evidence of deficits upon tests of phonemic and semantic fluency, or consistently identified one as being more impaired than the other. Thus, whilst some studies have reported deficits on both (Brown et al., 1994; Trichard et al., 1995), others have found no relationship between performance on these measures and the presence of depression (Austin et al., 1999; Crews, Harrison, & Rhodes, 1999; Feinstein, Goldberg, Nowlin, & Weinberger, 1998; Johnson & Crockett, 1982). Moreover, whilst some studies have reported that semantic fluency is more impaired than phonemic fluency (Calev, Nigal, & Chazan, 1989; Fossati et al., 1999; Geffen, Bate, Wright, Rozenbilds, & Geffen, 1993), the reverse (Beatty, Wonderlich, Staton, & Ternes, 1990), and a pattern of comparable impairment on the two measures has also been found (Elliott et al., 1996; King, Caine, Conwell, & Cox, 1991; Trichard et al., 1995).

To the present authors' knowledge, only two meta-analytic studies to date have quantified mean effects for both phonemic and semantic fluency, and in each it was found that semantic fluency was associated with a substantially larger deficit than phonemic fluency (Christensen et al., 1997; Zakzanis et al., 1998). However, whilst this would suggest that depression may be particularly associated with a degraded semantic store, in both meta-analytic reviews, the deficits upon both phonemic and semantic fluency were substantially in excess of the deficit for confrontation naming, a measure that, like semantic fluency, is also very sensitive to the integrity of semantic memory. Zakzanis et al. (1998) found that the effect sizes (expressed as Cohen's d) for semantic fluency,

phonemic fluency and confrontation naming were .97, .61 and .39 respectively; in Christensen et al.'s (1997) study, the corresponding values (expressed in a similar metric to Cohen's d, hedges g), were 1.07, .64 and .46. Since semantic fluency places relatively greater demands upon effortful processing than confrontation naming, it may only be when there is this added requirement that a prominent semantic memory deficit emerges. However, semantic fluency but not confrontation naming also places substantial demands upon cognitive speed. Thus, depression may be associated with a semantic memory deficit that is only apparent when there is additionally the requirement for speeded retrieval.

With respect to whether or not deficits on tests of phonemic and semantic fluency qualify as differential deficits relative to verbal intelligence (VIQ) and psychomotor speed, Zakzanis et al. (1998) and Christensen et al. (1997) found that deficits on phonemic and semantic fluency were both substantially in excess of deficits for VIQ. However, whilst Christensen et al. (1997) found that semantic but not phonemic fluency was associated with a larger deficit than measures of psychomotor speed (as indexed by the Trail Making Test Part A (TMT A) and Digit Symbol; Ms =1.07, .0.64, 0.90 and 0.97, respectively), Zakzanis et al. (1998) found that whilst both types of fluency were more impaired than the TMT A, only semantic fluency was more impaired than Digit Symbol (Ms = 0.97, 0.61, 0.18 and 0.63, respectively). Thus, although both meta-analyses suggest that the verbal fluency deficit in depression may not simply reflect a generalised verbal impairment, they have not agreed on whether these deficits are in excess of deficits upon measures of psychomotor speed.

Discrepancies between these meta-analyses may at least partially reflect the methodology that they employed, because in each there may have been substantive differences between the studies that assessed each of the constructs of interest, i.e., it is not the same studies that contribute

to each of the mean effect sizes being compared in each of these reviews. In Zakzanis et al.'s (1998) study, for instance, whilst eight studies contributed to the mean effect for VIQ, two contributed to the mean effect for semantic fluency; there may have been little or no overlap between the studies contributing to each of these statistics, and consequently little or no overlap in terms of the participants sampled. This raises a potentially important problem, particularly since it is probable that there are substantive differences between patients with depression, such as in the severity of the disorder. Thus, when assessing whether a deficit upon one measure is larger than a deficit upon another, it is important that the same participants contribute to both the measures of interest to 'control' for any potential differences between patients; as will be discussed, this methodology will be employed in the present analyses.

Thus, following on from these earlier analyses, in the present meta-analysis the relative prominence of deficits on tests of phonemic and semantic fluency will be investigated using a methodology that restricts the studies in each comparison of interest to only those that assess both the measures to be compared. This ensures that comparisons between different cognitive measures are fair as exactly the same participants will have been tested upon each measure. It should also be noted that in each of the previous meta-analytic reviews cited, fixed, as opposed to random effects meta-analytic models were employed for statistical analyses. As will be discussed, the National Research Council (1992) recommends use of the random effects model, and suggests that the fixed effects model should be the exception rather than the rule, as it may lead to inappropriately strong conclusions; in the present study, the random effects model will therefore be used. Finally, in both meta-analytic studies that have previously quantified mean effects for phonemic and semantic fluency, a relatively small number of effects contributed to each mean effect. In Zakzanis et al.'s (1998) study these means were calculated from seven and two effects, respectively, whilst for

Christensen et al.'s (1997) study these values were 14 and four, respectively; in the present study, over three times as many independent studies will contribute to each respective mean for phonemic and semantic fluency relative to previous meta-analytic reviews that have quantified these mean effects.

Discriminating Depression from DAT

Comparison of the relative magnitude of deficits on tests of phonemic and semantic fluency may also represent an effective means of discriminating between depression and dementia, disorders that are often associated with similar cognitive deficits, particularly in the early stages (McAllister & Price, 1982). Geffen et al. (1993) reported that, relative to controls, patients with depression and patients with DAT performed normally on a test of phonemic fluency but both were impaired on semantic fluency. However, DAT patients were significantly more impaired on semantic fluency relative to the depressed group ($p < .001$). Relatedly, King et al. (1991) found a significant difference between elderly depressed and DAT patients upon semantic, but not phonemic fluency. Thus, if the magnitude of the deficit for semantic fluency is greater in DAT than in depression, and the difference between phonemic and semantic effect sizes reliably larger in the former condition, this may aid in differential diagnosis. It is of note that whilst Christensen et al. (1997) compared the cognitive deficits associated with depression and DAT using meta-analytic techniques, the contrast between phonemic and semantic fluency in these two groups was not investigated.

Aims

The first aim (1) was to derive effect size estimates for phonemic and semantic fluency for patients with depression relative to healthy controls. Although other meta-analytic reviews have previously quantified these effect sizes, the present study will permit an extremely rigorous

cross-validation of their results, since, as noted, a substantially larger number of studies contribute to the present analyses. Moreover, as noted, in previous meta-analyses there may have been important differences in the studies contributing to each of these statistics. Therefore, in the present meta-analysis, in addition to calculating overall mean effects for phonemic and semantic fluency, the second aim (2) was to recalculate each of these mean effects, but with only studies that include both measures permitted to contribute. This will be the first meta-analysis to compare the magnitude of deficits on these measures whilst ensuring exactly the same studies contribute to each, and will therefore permit an extremely rigorous assessment of whether the verbal fluency deficit associated with depression predominantly reflects executive dysfunction, or problems with semantic memory, such as a specific semantic retrieval deficit (Henry & Crawford, in press-a).

However, as noted, the presence of a deficit on a test of phonemic or semantic fluency does not by itself provide evidence of executive or semantic memory dysfunction, respectively; it may instead, reflect a general verbal impairment, or psychomotor slowing. Thus, the third aim (3) was to estimate effect sizes for other cognitive measures in order to provide comparison standards, and thus assess to what extent fluency deficits in depression qualify as differential deficits. Premorbid intelligence as estimated by the National Adult Reading Test (NART; Nelson, 1982) or variants upon it, and the reading sub-test of the Wide Range Achievement Test (WRAT; Jastak & Wilkinson, 1984) was included to address the possibility that if a fluency deficit is present, it reflects the fact that patients with depression have not been successfully matched to their controls for premorbid ability. Also important was to address the possibility that phonemic and semantic fluency deficits simply reflect a generalised verbal dysfunction (see; Miller, 1984). Verbal fluency tests have a substantial verbal component, and indeed, phonemic fluency was originally developed as a measure of verbal intelligence (Thurstone & Thurstone, 1941). Thus, the pattern of deficits

across fluency versus verbal intelligence as measured by the WAIS (Wechsler, 1955, 1981) Verbal scale (VIQ) will be compared.

It will also be investigated whether deficits on tests of phonemic and semantic fluency are in excess of deficits on the WAIS Digit Symbol test (Wechsler, 1955, 1981) and Trail Making Test Part A (TMT A; Reitan, 1990), widely used measures of psychomotor speed. This will address the possibility that deficits on tests of verbal fluency simply reflect psychomotor retardation (i.e. a generalised reduction in cognitive speed), rather than executive dysfunction. Christensen et al. (1997) found that speeded relative to nonspeeded tasks were relatively more impaired. Thus, quantifying the influence of depression on measures of speed is important to address the possibility that deficits on tests designed to capture executive dysfunction simply reflect cognitive slowing. Performance on tests of fluency will also be compared with the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983) as confrontation naming is considered to be very sensitive to the integrity of semantic memory (Hart, 1988) yet imposes only minimal demands upon effortful retrieval and cognitive speed.

For episodic memory, the total and delayed scores will be recorded for verbal learning and delayed recall respectively from measures of the Auditory-Verbal Learning Test (Lezak, 1995) and California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987), with delayed recall from the Selective Reminding Test (Buschke & Fuld, 1974) also permitted to contribute to this latter construct. Quantifying mean effects for measures of this construct is particularly important given that another meta-analytic review found episodic, declarative memory to be associated with the greatest impairment in depression (Zakzanis et al., 1998). Finally, for executive functioning, performance on categories completed and perseverative errors on the WCST (WCST CC and WCST PE respectively; Heaton, 1981; Nelson, 1976) and the interference condition of the Stroop

(Golden, 1978) will be recorded. A comparison of phonemic and semantic fluency with these other putative measures of executive function will be extremely informative both with respect to their convergent validity, as well as their relative sensitivity to the presence of depression.

Finally, the fourth aim (4) is to compare the relative prominence of phonemic and semantic fluency deficits in depression with the corresponding statistics for patients with DAT. These comparisons are important because, as noted, depression and dementia are often associated with similar cognitive deficits, particularly in the early stages of each disorder (McAllister & Price, 1982). It has been suggested that semantic fluency is more sensitive to the presence of DAT than depression, but that phonemic fluency does not discriminate between the two groups (Geffen et al., 1993; King et al., 1991). This would predict that, relative to patients with depression, not only should patients with DAT present with a semantic fluency deficit that is larger, the difference between semantic and phonemic fluency should also be greater for this group. This possibility will be explored by comparing data from the present analyses with data taken from an independent meta-analysis involving patients with DAT (Henry & Crawford, submitted).

METHOD

Sample of Studies

A manual search of most issues of the Journal of Clinical and Experimental Neuropsychology Neuropsychology, Neuropsychologia, Cognitive Neuropsychiatry, The Clinical Neuropsychologist, The Journal of the International Neuropsychological Society, Neuropsychiatry, Neuropsychology and Behavioural Neurology, and the Journal of Neuropsychiatry and Clinical Neurosciences was conducted. A computer-based search involving the Web of Science, Psych Lit CD-ROM, and Science Direct databases was also undertaken, using the following terms as search parameters; 'letter fluency',

'FAS', 'semantic fluency', 'category fluency', 'controlled oral word association', 'COWA(T)', 'word fluency' 'verbal fluency', 'oral fluency', 'phonemic fluency', 'executive test' and 'frontal test'. The search was completed in December 2002.

The inclusion criteria were (1) the patient group had to consist entirely of adults that met accepted criteria for primary depression such as Diagnostic and Statistical Manual of Mental Disorders criteria (DSM; American Psychiatric Association, 1994) although not necessarily major depression. Christensen et al. (1997) found that the type of depression did not significantly influence the magnitude of the mean cognitive deficit. Thus, patients with, for instance, dysthymia and bipolar disorders were included, although bipolar patients were only eligible when they were tested in the depressed stage of the illness. In addition, the study had to include (2) a healthy control group free from neurological or psychiatric disease, and (3) a measure of phonemic and/or semantic fluency. Effect size estimates for premorbid intelligence, VIQ, TMT A, Digit Symbol, BNT, Verbal Learning, Delayed Recall, WCST CC, WCST PE and Stroop were derived from studies that also reported verbal fluency results. For inclusion, the study must also have (4) presented precise statistics convertible to effect size r, (5) been published, (6) in English, (7) in a journal.

Statistical Analysis

Meta-analysis is a rigorous, quantitative alternative to the traditional review process, as it involves statistical integration of results. The basis of this methodology is the effect size, a standardised statistic that quantifies the magnitude of an effect. In the present study the effect size r was employed, which corresponds to the degree of correlation between group membership (i.e. the presence or absence of depression), and performance on the cognitive measure of interest. It should be noted that because the correlation coefficient is associated with a slight bias, Fisher (1928) derived a

transformation of r that Snedecor (1989) has recommended should be employed during statistical analyses in preference to r. However, this transformed estimate is itself associated with a bias, and in a Monte Carlo analysis Field (2001) reported that for random effects meta-analytic models, transformed effect-size estimates produced substantial upward biases of a larger magnitude than the corresponding downward biases associated with untransformed correlation coefficients. Thus, in the present study, untransformed correlation coefficients have been employed for statistical analyses.

For each construct, effects were pooled to derive an estimate of the mean, with each effect weighted for sample size to correct for sampling error. To do so, the random effects meta-analytic model (Shadish & Haddock, 1994) was selected in preference to the more commonly employed fixed effects model as it yields more generalisable parameter estimates. This is because, in the fixed effects model, the mean is presumed to reflect a common underlying effect parameter that gives rise to the sample observations. However, in the random effects model the mean represents a hyperparameter, as it allows for substantive differences beyond sampling error that differentiate the effects contributing to each respective mean (Raudenbush, 1994).

Statistically, the crucial difference between these methodologies is in the calculation of standard errors and confidence intervals, which for the random effects model are typically larger. The National Research Council (1992) argues that the fixed effects model should be the exception rather than the rule, as it may lead to inappropriately strong conclusions. Thus, although more technically demanding, it was considered important to use the random effects model in the present work.

To estimate the degree of heterogeneity of the effects contributing to each mean, the homogeneity statistic Q and the random effects variance (σ_θ^2) were estimated, as well as the SD of random effects, and the 95% confidence intervals (CI) within which random effects can be expected to fall. Q quantifies within-group heterogeneity (i.e. the degree to which the studies contributing to each

respective mean can be regarded as homogenous). If the Q statistic associated with a mean effect is significant, this suggests that there are substantive differences between the studies contributing to that particular mean. In contrast, a non-significant estimate of Q suggests that once sampling error has been removed, no substantive differences between the studies contributing to the respective mean in question remain (i.e. the null hypothesis of homogeneity of effects cannot be rejected).

As noted earlier, since there may be substantive differences between patients with depression such as the severity of the disorder, for each comparison of interest, only studies that assessed both variables of interest were included. For example, although in total 53 depressed groups were assessed upon phonemic fluency, and 15 groups upon semantic fluency, since only 14 depressed groups were assessed on both phonemic and semantic fluency, when comparing the magnitude of deficits upon tests of phonemic and semantic fluency, only these 14 groups were permitted to contribute to the analyses. This effectively 'controls' for substantive differences between studies when comparing two different cognitive measures, as it ensures that exactly the same participants are being compared upon each of these measures.

Mean effects were also calculated for each of the non-fluency variables identified (premorbid IQ, VIQ, TMT A, Digit Symbol, BNT, Verbal Learning, Delayed Recall, WCST CC, WCST PE and Stroop) and compared with phonemic and, where enough studies were available, semantic fluency. Again, to control for any potential differences between studies, only studies that assessed both the fluency and the particular non-fluency variable of interest were included in each comparison. Thus, for each comparison the mean effect for phonemic or semantic fluency was re-calculated using only those studies that assessed both the fluency and non-fluency measures of interest.

Finally, the null hypothesis that the mean effect size is zero was tested with the statistic Z; if the value of Z exceeds 1.96, this indicates that the mean effect differs significantly from zero at the .05

level. To interpret how important a particular effect was in practical terms, Cohen's (1977) guidelines were adopted. These suggest that a correlation of 0.1 should be regarded as representing a small effect, 0.3 as medium, and 0.5 as large. In addition, squares of the effect size multiplied by 100 were also presented as these latter quantities represent the percentage of the variance accounted for (PVAF) by group membership (i.e. the presence of depression versus being a member of the healthy adult population) on a measure of interest.

RESULTS

Participant Characteristics

Forty two studies published between 1982 and 2002 met the inclusion criteria, in which there was a total of 1077 patients with depression and 1129 controls. These studies are marked by an asterisk in the reference section. Patients and controls were closely matched for age (M = 52.7, SD =13.69 versus M = 50.8, SD = 16.19 respectively), education (M = 12.8, SD = 1.80 versus M = 13.2, SD = 1.88, respectively), and the proportion of the patient group that was male (39.0% versus 38.8%, respectively).

Verbal Fluency Performance for Patients with Depression

Study-level effect sizes for phonemic and semantic fluency are presented in Appendix A. A positive sign indicates that patients have performed worse than controls, a negative sign the reverse.

[TABLE 1 ABOUT HERE]

Table 1 presents estimates of the mean effects, their variability, and practical importance for measures of phonemic and semantic fluency. Whereas K refers to the number of groups of patients with depression contributing to each respective statistic, N refers to the number of participants (i.e. the

total number of patients plus controls). Both mean effects are significantly different from zero (ps < .05), and moderate in magnitude. However, for semantic fluency the mean effect (r = .44) is substantially larger than for phonemic fluency (r = .30). For both mean effects estimates of Q are significant (ps < .05), reflecting the heterogeneous nature of the disorder.

Phonemic and semantic fluency deficits relative to other cognitive deficits

Table 2 presents estimates of the mean effects for phonemic and semantic fluency, their variability, and practical importance for studies that include both of these measures. In addition, mean effects are presented for premorbid IQ, VIQ, TMT A, Digit Symbol, BNT, Verbal Learning, Delayed Recall, WCST CC, WCST PE and Stroop interference, and were calculated using only those studies that included the particular non-fluency measures of interest in addition to either phonemic or semantic fluency. As noted previously, this methodology ensures that exactly the same participants are contributing to the mean effects for the two variables being compared.

[TABLE 2 ABOUT HERE]

Thus, it can be seen in Table 2 that for each non-fluency measure, for instance premorbid IQ, two mean effects have been calculated; one for studies that also assess phonemic fluency (r = .19; K = 17), and one for studies that also assess semantic fluency (r = .27, K = 7). Each fluency mean effect was also re-calculated for each of these comparisons; these data are presented in the final column of Table 2. For comparisons with semantic fluency, it should be noted that using this more rigorous methodology it was only possible to compare performance on this variable with phonemic fluency, premorbid IQ and VIQ, as very few studies presented the prerequisite data for the other variables of interest in addition to semantic fluency.

For studies that assessed both phonemic and semantic fluency, the magnitude of deficits upon each were very highly correlated (r = .77, p = .001). It can be seen in Table 2 that, in contrast to the results presented in Table 1 which demonstrated a substantially larger deficit on semantic relative to phonemic fluency, when only studies that assess both these measures are included in analyses, the deficit for semantic fluency is only marginally larger than the deficit for phonemic fluency (rs = .43 versus .39).

The effect size for phonemic fluency is in excess of the effect sizes for premorbid and current verbal intelligence (rs = .28 versus .19 for phonemic fluency versus premorbid IQ; rs = .26 versus .17 for phonemic fluency versus VIQ); this is also true for comparisons of semantic fluency with premorbid and current intelligence.

As noted, for comparisons with the other variables of interest, because of the methodology used comparisons were only possible with phonemic fluency. It can be seen that relative to measures of psychomotor speed (Digit Symbol and TMT A), the phonemic fluency deficit is of a smaller or comparable magnitude. The deficit for the BNT is not even small in magnitude, and substantially smaller than the corresponding mean effect for phonemic fluency (rs = .05 versus .20). However, measures of episodic memory (verbal learning and delayed recall) are associated with larger deficits than phonemic fluency. Finally, relative to other measures of executive functioning (WCST CC, WCST PE and Stroop), phonemic fluency deficits are of a comparable magnitude.

Depression Compared to DAT

[TABLE 3 ABOUT HERE]

It has been reported that semantic fluency is more impaired in DAT than depression, but that phonemic fluency performance does not differentiate the two groups (Geffen et al., 1993; King et

al., 1991). In Table 3 mean effects for phonemic and semantic fluency are stratified according to diagnostic group (Depression and DAT; data for patients with DAT taken from; Henry and Crawford, submitted). It should be noted that for the DAT analyses, the same 70 studies contribute to the statistics for both phonemic and semantic fluency.

It can be seen that patients with DAT are more impaired on both phonemic and semantic fluency relative to patients with depression, but that for both groups semantic fluency is more impaired than phonemic fluency. However, the difference in terms of the PVAF between the fluency measures is substantially larger for patients with DAT compared to patients with depression (Δ PVAF = 20.80 and 3.28% respectively).

Assessing the Potential Presence of Publication Bias

A number of validity threats have been identified that may lead to imprecise conclusions in both non-quantitative and meta-analytic reviews. Particularly problematic is 'the file drawer problem' which refers to the fact that significant results are more likely to be published than non-significant results. To assess whether this bias posed a threat, funnel plot diagrams were constructed in which sample size is plotted against the corresponding study-level effect. If statistically non-significant results have been discriminated against, there should be a relative absence of studies with small sample sizes that report weak effects. For none of the variables was there evidence of this bias operating.

DISCUSSION

When mean effects were calculated for phonemic and semantic fluency including all studies that had assessed either of these measures, it was found that the semantic fluency deficit was the

substantially larger of the two (rs = .44 versus .30). It is of interest that the magnitude of these deficits are very much in agreement with the results of previous meta-analytic reviews that have quantified mean effects for each of these measures. Expressed as Cohen's d, a different metric of effect size, the semantic and phonemic fluency deficits in the current study were found to be .98 and .63; in Zakzanis et al.'s (1998) meta-analysis, the corresponding values were .97 and .61, and in Christensen et al.'s (1997) meta-analysis (expressed in a similar metric to Cohen's d, hedges g), these values were 1.07 and .64. Since it is thought that, relative to phonemic fluency, semantic fluency imposes greater demands upon the integrity of the semantic network, but the two measures are equivalent in sensitivity to executive control processes such as effortful retrieval (Henry & Crawford, in press-a), a substantially larger deficit for semantic relative to phonemic fluency suggests that the semantic memory deficit associated with depression reflects a degradation or disorganisation of the semantic store, and not retrieval slowing (see also; Hodges, Salmon, & Butters, 1992; Salmon, Heindel, & Lange, 1999).

However, as noted earlier, it is important that the patients contributing to the mean effect for semantic fluency do not differ from the patients contributing to the mean for phonemic fluency, if comparisons between these two measures are to be fair. This rules out, for instance, the possibility that the group of patients contributing to the mean for the former are more severely depressed than those contributing to the mean for the latter. When mean effects were re-calculated in the present study using a more rigorous methodology based only on studies that had assessed both phonemic and semantic fluency, it was found that there is in fact very little difference in the relative sensitivity of the two measures to the presence of depression (rs = .43 and .39 for semantic and phonemic fluency, respectively). Moreover, for studies that assessed both phonemic fluency and the BNT, the former was not only substantially larger; the latter was of only a negligible magnitude (rs = .20

versus .05, respectively). Thus, when a more rigorous methodology is applied, the pattern of deficits across phonemic fluency, semantic fluency and the BNT is more classically 'frontal' than 'temporal' (Henry & Crawford, in press-a), and provides no evidence that the disorder is associated with a degraded or disrupted semantic store.

Evidence for a Differential Executive Deficit?

As discussed previously, a pattern of comparable impairment upon tests of phonemic and semantic fluency for patients with depression may reflect executive dysfunction if, as was found for frontal patients (but not for non-frontal patients; Henry & Crawford, in press-a), verbal fluency deficits qualify as differential deficits relative to verbal intelligence and psychomotor speed. In the present study, this pattern of results was not found. Although the fluency deficits did not simply reflect a failure to match patients and controls upon premorbid verbal IQ, and relative to current VIQ both types of fluency were more impaired, the phonemic fluency deficit was of a smaller or comparable magnitude to deficits on tests of psychomotor speed (Digit Symbol and the TMT A). Thus, cognitive slowing may at least partially underlie deficits on these and possibly other measures of executive function. As noted earlier, it was found in another meta-analytic review that speeded relative to non-speeded tasks were substantially more impaired, also consistent with the possibility that the cognitive dysfunction associated with depressive illness may reflect cognitive slowing (Christensen et al., 1997). Indeed, it is noteworthy that in the present study the smallest deficit identified was associated with a measure of semantic memory that imposes only minimal demands upon cognitive speed, the BNT ($\underline{r}$ = .05).

However, it has also been suggested that a potential limitation to the interpretation of any cognitive deficits in depression (as well as a number of other cognitive disorders) is that the level of

effort is not controlled for in most studies. The importance of this variable as a potential moderator of effects has been highlighted by work by Rohling, Green, Allen and Iverson (2002). In this study two symptom validity measures were administered to consecutive referrals with heterogeneous diagnoses in order to exclude patients who were not demonstrating satisfactory effort. Two sub-groups were then derived from the remaining 420 patients, based on level of depression as indexed by the Beck Depression Inventory (Beck, 1978). Whilst those in the lowest quartile were assigned to the low depression group (N = 115), those in the top quartile were allocated to the high depression group (N = 112). It was found that in this sample of patients in which patients who had performed sub-optimally were excluded, the presence of depression had no effect on patients' performance on a battery of objective cognitive and psychomotor tests. Rohling et al. (2002) suggested that when patients with depression perform poorly on tests of cognitive ability, this may therefore not reflect the presence of depression, but rather sub-optimal effort when undergoing testing.

However, irrespective of whether speed or sub-optimal effort is the critical determinant of the magnitude of the cognitive deficits observed, the results of the present study suggest that the role of executive dysfunction in depression may have been over-estimated, particularly since in neuropsychological research phonemic fluency represents one of the most widely used measures of this construct, and evidence for its construct validity is particularly strong. It is also of interest that compared with other non-fluency measures of executive function, phonemic fluency was associated with very similar effect size estimates. Relative to the WCST, phonemic fluency has proven to be far more sensitive to both the presence of focal cortical lesions (Henry & Crawford, in press-a), and traumatic brain injury (Henry & Crawford, in press-b; Azouvi, Jokic, Van Der Linden, Marlier, & Bussel, 1996), yet the present results suggest that these measures are equivalent in sensitivity to

depression. As noted earlier, there is evidence for at least three dissociable executive processes (Miyake et al., 2000). Despite the differential reliance of these executive measures upon different component processes, the present study could not differentiate between them in terms of their relative sensitivity to depression. Thus, the present results provide no evidence that each of these cognitive control processes are differentially affected by the presence of this disorder.

Finally, the only measures to be associated with a substantially larger deficit than phonemic fluency were verbal learning and delayed recall. As noted, in another meta-analytic study it was found that deficits in episodic memory were the most prominent disturbances associated with depressive illness (Zakzanis et al., 1998). The present study provides additional evidence that depression is associated with deficits in episodic memory, and is consistent with Zakzanis et al.'s (1998) study in that it suggests that these deficits may be more pronounced than executive dysfunction.

Future Directions

The mean effects for virtually all of the cognitive measures assessed in the present study were associated with significant heterogeneity. Since sampling error, the most serious source of artefactual variance, had been removed, substantive differences between studies remain. Some of this variance will almost certainly be attributable to differences in severity between patients. However, it is important to stress that depression is an extremely heterogeneous disorder, and it is therefore likely that other factors in addition to illness severity will also moderate effect size magnitude upon tests of fluency as well as other cognitive measures across individual studies. The present results are therefore consistent with the possibility that there are a number of important moderating factors that result in the heterogeneity observed which may include melancholic versus

non-melancholic depression (Austin et al., 1999), the presence or absence of white matter hyperintensities (Kramer-Ginsberg et al., 1999), as well as the severity of the depression, age of onset and duration of the illness.

Thus, whilst in general patients with depression can be expected to be comparably impaired on semantic and phonemic fluency, for certain sub-groups this may not be the case. Indeed, as noted previously, some studies have reported that semantic fluency is more impaired than phonemic fluency (Calev et al., 1989; Fossati et al., 1999; Geffen et al., 1993), and others the reverse (Beatty et al., 1990). Thus, the present results do not rule out the possibility that distinct sub-types may exist for whom executive dysfunction or semantic memory dysfunction is particularly prominent. McKenna (1991), for instance, has suggested that delusions are associated with an inappropriate development of new semantic memories. Patients with depression who are psychotic or deluded may therefore be expected to have dysfunctional semantic memory systems. Indeed, it is of note that Rossell et al. (1999) found that schizophrenic patients with delusions were significantly more impaired on semantic fluency than schizophrenic patients without delusions, and this was hypothesised to reflect a preoccupation with delusional ideas strengthening associations between unrelated concepts.

It is unfortunate that much of this variance will be bundled up within, rather than between studies, and thus the specific influence of each cannot be explored in the present study (for example, there were insufficient studies that were restricted only to patients with delusions versus those without, etc.). The statistic Q quantifies the degree of heterogeneity between studies but cannot address the degree of heterogeneity within each of the studies contributing to a mean. However, it is recommended that if future primary research breaks down their samples more fully, meta-analysis should be conducted to address which of these variables moderate performance on tests of verbal fluency.

Verbal Fluency in Depression Compared with DAT

Making an accurate and early diagnosis of either major depression or DAT is extremely important both in terms of prognosis, and ensuring that the correct treatment is administered. However, it is recognised that there is often a great deal of overlap in the clinical symptoms of major depression and DAT, such as apathy, social inhibition and depressed mood. Moreover, depression and DAT are often associated with similar cognitive deficits in the early stages (McAllister & Price, 1982), and to date no accurate biological marker has been identified that distinguishes between them. Thus, differentiating these two disorders is considered to be one of the most challenging diagnostic issues in psychiatry.

The present study found that patients with early DAT are more impaired on both phonemic and semantic fluency relative to patients with depression, and that for both groups semantic fluency was more impaired than phonemic fluency. However, the difference in terms of the PVAF between phonemic and semantic fluency was substantially larger for patients with DAT than for those with depression (Δ PVAF = 20.80% and 3.28%, respectively). This is consistent with previous studies that have investigated performance on tests of fluency in these two groups (Geffen et al., 1993; King et al., 1991), and suggests that, relative to depression, DAT is typically associated with greater disruption to the semantic system.

It is of note that in their meta-analysis of DAT patients Henry and Crawford (submitted) found that although the effect sizes for both phonemic and semantic fluency were significantly positively related to dementia severity, the difference in terms of the PVAF by these measures was not (r = .16, p = .29). This suggests that deficits in semantic memory storage exceed executive dysfunction to a comparable degree at every stage of DAT. Thus, even in the mild-moderate stages

of DAT, when differentiating this disorder from major depression has proven to be particularly difficult, the present results suggest that comparison of the relative magnitude of the deficits on tests of semantic and phonemic fluency may prove useful in supporting the differential diagnosis of DAT and major depression.

Summary and Conclusions

Contrary to the findings from previous meta-analytic reviews, the present study provides strong evidence that phonemic and semantic fluency are broadly equivalent in their sensitivity to the presence of depression. Although the effect sizes associated with both of these measures were in excess of those for premorbid and current verbal intelligence, neither of these deficits qualified as differential deficits relative to measures of psychomotor speed. Thus, patients with depression do not appear to perform poorly on these tasks as a consequence of executive dysfunction or a degraded semantic store, but instead because of a more generalised deficit, such as cognitive slowing. Differential diagnosis of patients with depression and those in the early stages of DAT may be aided by including tests of phonemic and semantic fluency in the assessment battery.

REFERENCES

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual Review of Neuroscience, 9, 357-381.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: American Psychiatric Association.

Austin, M. P., Mitchell, P., Wilhelm, K., Parker, G., Hickie, I., Brodaty, H., Chan, J., Eyers, K., Milic, M., & Hadzi-Pavlovic, D. (1999). Cognitive function in depression: a distinct pattern of frontal impairment in melancholia? Psychological Medicine, 29, 73-85.*

Austin, M. P., Ross, M., Murray, C., O'Carroll, R. E., Ebmeier, K. P., & Goodwin, G. M. (1992). Cognitive function in major depression. Journal of Affective Disorders, 25, 21-30.*

Azouvi, P., Jokic, C., Van Der Linden, M., Marlier, N., & Bussel, B. (1996). Working memory and supervisory control after severe closed-head injury. A study of dual task performance and random generation. Journal of Clinical and Experimental Neuropsychology, 18, 317-337.

Basso, M. R., Lowery, N., Neel, J., Purdie, R., & Bornstein, R. A. (2002). Neuropsychological impairment among manic, depressed, and mixed-episode inpatients with bipolar disorder. Neuropsychology, 16, 84-91.*

Beatty, W. W., Wonderlich, S. A., Staton, R. D., & Ternes, L. A. (1990). Cognitive functioning in bulimia: Comparison with depression. Bulletin of the Psychonomic Society, 28, 289-292.*

Beck, A. T. (1978). Beck depression inventory. San Antonio, TX: The Psychological Corporation Manual.

Blackwood, S. K., MacHale, S. M., Power, M. J., Goodwin, G. M., & Lawrie, S. M. (1998). Effects of exercise on cognitive and motor function in chronic fatigue syndrome and depression. Journal of Neurology Neurosurgery and Psychiatry, 65, 541-546.*

Boone, K. B., Lesser, I., Miller, B., Wohl, M., Berman, N., Lee, A., & Palmer, B. (1994). Cognitive functioning in a mildly to moderately depressed geriatric sample: relationship to chronological age. The Journal of Neuropsychiatry and Clinical Neurosciences, 6, 267-272.*

Boone, K. B., Lesser, I. M., Miller, B. L., Wohl, M., Berman, N., Lee, A., Palmer, B., & Back, C. (1995). Cognitive functioning in older depressed outpatients: relationship of presence and severity of depression to neuropsychological test scores. Neuropsychology, 9, 390-398.*

Brown, R. G., Scott, L. C., Bench, C. J., & Dolan, R. J. (1994). Cognitive function in depression: its relationship to the presence and severity of intellectual decline. Psychological Medicine, 24, 829-847.*

Buschke, H., & Fuld, P. A. (1974). Evaluation of storage, retention, and retrieval in disordered memory and learning. Neurology, 11, 1019-1025.

Calev, A., Nigal, D., & Chazan, S. (1989). Retrieval from semantic memory using meaningful and meaningless constructs by depressed, stable bipolar and manic patients. British Journal of Clinical Psychology, 28, 67-73.

Christensen, H., Griffiths, K., Mackinnon, A., & Jacomb, P. (1997). A quantitative review of cognitive deficits in depression and Alzheimer-type dementia. Journal of the International Neuropsychological Society, 3, 631-651.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (Revised ed.). New York: Academic Press.

Conaghan, S., & Davidson, K. M. (2002). Hopelessness and the anticipation of positive and negative future experiences in older parasuicidal adults. British Journal of Clinical Psychology, 41, 233-242.*

Crawford, J. R., & Henry, J. D. (in press). Assessment of executive deficits. In P. W. Halligan & N. Wade (Eds.), The effectiveness of rehabilitation for cognitive deficits. London: Oxford University Press.

Crews, W. D., Harrison, D. W., & Rhodes, R. D. (1999). Neuropsychological test performances of young depressed outpatient women: An examination of executive functions. Archives of Clinical Neuropsychology, 14, 517-529.*

Crowe, S. F. (1996). The performance of schizophrenic and depressed subjects on tests of fluency: Support for a compromise in dorsolateral prefrontal functioning. Australian Psychologist, 31, 204-209.*

Crowe, S. F., & Hoogenraad, K. (2000). Differentiation of dementia of the Alzheimer's type from depression with cognitive impairment on the basis of a cortical versus subcortical pattern of cognitive deficit. Archives of Clinical Neuropsychology, 15, 9-19.*

David, A. S. (1992). Frontal lobology - psychiatry's new pseudoscience. British Journal of Psychiatry, 161, 244-248.

de Groot, M. H., Nolen, W. A., Huijsman, A. M., & Bouvy, P. F. (1996). Lateralized neuropsychological functioning in depressive patients before and after drug therapy. Biological Psychiatry, 40, 1282-1287.*

Degl'Innocenti, A., Agren, H., & Backman, L. (1998). Executive deficits in major depression. Acta Psychiatrica Scandinavica, 97, 182-188.*

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). California Verbal learning Test adult version. New York: The Psychological Corporation.

Della Sala, S., Gray, C., Spinnler, H., & Trivelli, C. (1998). Frontal lobe functioning in man: The riddle revisited. Archives of Clinical Neuropsychology, 13, 663-682.

Elliott, R., Sahakian, B. J., McKay, A. P., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1996). Neuropsychological impairments in unipolar depression: The influence of perceived failure on subsequent performance. Psychological Medicine, 26, 975-989.*

Evangeli, M., & Broks, P. (2000). Face processing in schizophrenia: Parallels with the effects of amygdala damage. Cognitive Neuropsychiatry, 5, 81-104.*

Feinstein, A., Goldberg, T. E., Nowlin, B., & Weinberger, D. R. (1998). Types and characteristics of remote memory impairment in schizophrenia. Schizophrenia Research, 30, 155-163.*

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects models. Psychological Methods, 6, 161-180.

Fisher, R. A. (1928). Statistical methods for research workers. (Second ed.). London: Oliver & Boyd.

Flint, A. J., & Eastwood, M. R. (1988). Frontal lobe syndrome and depression in old age. Journal of Geriatric Psychiatry and Neurology, 1, 53-55.

Fossati, P., Amar, G., Raoux, N., Ergis, A. M., & Allilaire, J. F. (1999). Executive functioning and verbal memory in young patients with unipolar depression and schizophrenia. Psychiatry Research, 89, 171-187.*

Fossati, P., Ergis, A. M., & Allilaire, J. F. (2002). Executive functioning in unipolar depression: a review. Encephale-Revue de psychiatre cliniique biologique et therapeutique, 28, 97-107.

Franke, P., Maier, W., Hardt, J., Frieboes, R., Lichtermann, D., & Hain, C. (1993). Assessment of frontal lobe functioning in schizophrenia and unipolar major depression. Psychopathology, 26, 76-84.*

Gaudino, E. A., Masur, D. M., Kaufman, L. D., Sliwinski, M., & Krupp, L. B. (1995). Depression and neuropsychological performance in the eosinophilia myalgia syndrome: a comprehensive analysis of cognitive function in a chronic illness. Neuropsychiatry Neuropsychology and Behavioral Neurology, 8, 118-126.*

Geffen, G., Bate, A., Wright, M., Rozenbilds, U., & Geffen, L. (1993). A comparison of cognitive impairments in dementia of the Alzheimer type and depression in the elderly. Dementia, 4, 294-300.*

Golden, C. J. (1978). Stroop Color and Word Test. Chicago, IL: Stoelting.

Hart, R. P., Kwentus, J. A., Taylor, J. R., & Hamer, R. M. (1988). Productive naming and memory in depression and Alzheimer's type dementia. Archives of Clinical Neuropsychology, 3, 313-322.*

Hart, R. P., Kwentus, J. A., Taylor, J. R., & Harkins, S. W. (1987). Rate of forgetting in dementia and depression. Journal of Consulting and Clinical Psychology, 55, 101-105.*

Hart, S. (1988). Language and Dementia: a review. Psychological Medicine, 18, 99-112.

Heaton, R. K. (1981). Wisconsin Card Sorting Test (WCST). Odessa, FL: Psychological Assessment Resources.

Henry, J. D., & Crawford, J. R. (in press-a). A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology.

Henry, J. D., & Crawford, J. R. (in press-b). A meta-analytic review of verbal fluency performance in traumatic brain injured patients. Neuropsychology.

Henry, J. D., & Crawford, J. R. (submitted). Verbal fluency performance in dementia of the Alzheimer's type; A meta-analysis.

Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease - failure of access or degraded knowledge. Neuropsychologia, 30, 301-314.

Ilsley, J. E., Moffoot, A. P. R., & O'Carroll, R. E. (1995). An analysis of memory dysfunction in major depression. Journal of Affective Disorders, 35, 1-9.*

Jastak, S., & Wilkinson, G. S. (1984). Wide Range Achievement Test-Revised. Wilmington, DE: Jastak Assessment Systems.

Johnson, O., & Crockett, D. (1982). Changes in perceptual asymmetries with clinical improvement of depression and schizophrenia. Journal of Abnormal Psychology, 91, 45-54.*

Kaiser, S., Unger, J., Kiefer, M., Markela, J., Mundt, C., & Weisbrod, M. (2003). Executive control deficit in depression: event-related potentials in a Go/Nogo task. Psychiatry Research: Neuroimaging, 122, 169-184.

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). The Boston Naming Test. Philadelphia: Lea & Febiger.

King, D. A., Caine, E. D., Conwell, Y., & Cox, C. (1991). The neuropsychology of depression in the elderly: a comparative study of normal aging and Alzheimer's disease. The Journal of Neuropsychiatry and Clinical Neurosciences, 3, 163-168.*

King, D. A., Cox, C., Lyness, J. M., Conwell, Y., & Caine, E. D. (1998). Quantitative and qualitative differences in the verbal learning performance of elderly depressives and healthy controls. Journal of the International Neuropsychological Society, 4, 115-126.*

Kramer-Ginsberg, E., Greenwald, B. S., Krishnan, K. R. R., Christiansen, B., Hu, J., Ashtari, M., Patel, M., & Pollack, S. (1999). Neuropsychological functioning and MRI signal hyperintensities in geriatric depression. American Journal of Psychiatry, 156, 438-444.*

Kuzis, G., Sabe, L., Tiberti, C., Leiguarda, R., & Starkstein, S. E. (1997). Cognitive functions in major depression and Parkinson disease. Archives of Neurology, 54, 982-986.*

Lafont, V., Medecin, I., Robert, P. H., Beaulieu, F. E., Kazes, M., Danion, J. M., Pringuey, D., & Darcourt, G. (1998). Initiation and supervisory processes in schizophrenia and depression. Schizophrenia Research, 34, 49-57.*

Landro, N. I., Stiles, T. C., & Sletvold, H. (1997). Memory functioning in patients with primary fibromyalgia and major depression and healthy controls. Journal of Psychosomatic Research, 42, 297-306.*

Landro, N. I., Stiles, T. C., & Sletvold, H. (2001). Neuropsychological function in nonpsychotic unipolar major depression. Neuropsychiatry, Neuropsychology and Behavioural Neurology, 14, 233-240.*

Laws, K. R. (1999). A meta-analytic review of wisconsin card sort studies in schizophrenia: general intellectual deficit in disguise? Cognitive Neuropsychiatry, 4, 1-35.

Lesser, I. M., Miller, B. L., Boone, K. B., Hill-Gutierrez, E., Mehringer, C. M., Wong, K., & Mena, I. (1991). Brain injury and cognitive function in late-onset psychotic depression. The Journal of Neuropsychiatry and Clinical Neurosciences, 3, 33-40.*

Lezak, M. D. (1995). Neuropsychological Assessment. (Third ed.). New York: Oxford University Press.

Lyness, S. A., Eaton, E. M., & Schneider, L. S. (1994). Cognitive performance in older and middle-aged depressed outpatients and controls. Journals of Gerontology, 49, 129-136.*

MacLeod, A. K., & Salaminiou, E. (2001). Reduced positive future-thinking in depression: cognitive and affective factors. Cognition and Emotion, 15, 99-107.*

MacLeod, A. K., Tata, P., Kentish, J., & Jacobsen, H. (1997). Retrospective and prospective cognitions in anxiety and depression. Cognition and Emotion, 11, 467-479.*

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain; structure and processes. Current Opinion in Neurobiology, 11, 194-201.

McAllister, T. W., & Price, T. R. (1982). Severe depressive pseudodementia with and without dementia. American Journal of Psychiatry, 139, 626-629.

McKenna, P. J. (1991). Memory, knowledge and delusions. British Journal of Psychiatry, 159, 36-41.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. British Journal of Clinical Psychology, 23, 53-57.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cognitive Psychology, 41, 49-100.

National Research Council (1992). Combining information; statistical issues and opportunities for research. Washington DC: National Academy Press.

Nathan, J., Wilkinson, D., Stammers, S., & Low, J. L. (2001). The role of tests of frontal executive function in the detection of mild dementia. International Journal of Geriatric Psychiatry, 16, 18-26.*

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. Cortex, 12, 313-324.

Nelson, H. E. (1982). National Adult Reading Test (NART): Test Manual. Windsor, UK: NFER Nelson.

Norris, M. P., Blankenship-Reuter, L., Snow-Turek, A. L., & Finch, J. (1995). Influence of depression on verbal fluency performance. Aging and Cognition, 2, 206-215.

Pendleton Jones, B., Henderson, M., & Welch, C. A. (1988). Executive functions in unipolar depression before and after electroconvulsive therapy. International Journal of Neuroscience, 38, 287-297.*

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. Neuropsychologia, 12, 323-330.

Phillips, L. H. (1997). Do 'frontal tests' measure executive function? Issues of assessment and evidence from fluency tests. In P. M. A. Rabbitt (Ed.), Methodology of Frontal and Executive Function (pp. 191-213): Hove: Psychology Press.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), The Handbook of Research Synthesis (pp. 301-321). New York: Russell Sage Foundation.

Reitan, W. (1990). The Halstead-Reitan Neuropsychological Test Battery, theory and clinical interpretation.: Neuropsychology Press.

Rogers, M. A., Bradshaw, J. L., Pantellis, C., & Phillips, J. G. (1998). Frontostriatal deficits in unipolar major depression. Brain Research Bulletin, 47, 297-310.

Rohrer, D., Salmon, D. P., Wixted, J. T., & Paulsen, J. S. (1999). The disparate effects of Alzheimer's disease and Huntington's disease on semantic memory. Neuropsychology, 13, 381-388.

Rohling, M. L., Green, P., Allen, L. M., & Iverson, G. L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. Archives of Clinical Neuropsychology, 17, 205-222.

Rossell, S. L., Rabe-Hesketh, S., Shapleske, J., & David, A. S. (1999). Is semantic fluency differentially impaired in schizophrenic patients with delusions? Journal of Clinical and Experimental Neuropsychology, 21, 629-642.

Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1997). The psychological construct of word fluency. Brain and Language, 57, 394-405.

Salmon, D. P., Heindel, W. C., & Lange, K. L. (1999). Differential decline in word generation from phonemic and semantic categories during the course of Alzheimer's disease: Implications for the integrity of semantic memory. Journal of the International Neuropsychological Society, 5, 692-703.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), The Handbook of Research Synthesis (pp. 261-281). New York: Russell Sage Foundation.

Shah, P. J., Ogilvie, A. D., Goodwin, G. M., & Ebmeier, K. P. (1997). Clinical and psychometric correlates of dopamine D-2 binding in depression. Psychological Medicine, 27, 1247-1256.*

Shallice, T. (1988). From neuropsychology to mental structure. Cambridge: Cambridge University Press.

Smith, M. J., Brebion, G., Banquet, J. P., & Allilaire, J. F. (1994). Experimental evidence for two dimensions of cognitive disorders in depressives. Journal of Psychiatric Research, 28, 401-411.*

Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods (Eighth ed.): Iowa State University Press.

Stuss, D. T., & Benson, D. F. (1986). The frontal lobes. New York: Raven Press.

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. Psychometric Monographs, 2, 94.

Trichard, C., Martinot, J. L., Alagille, M., Masure, M. C., Hardy, P., Ginestet, D., & Feline, A. (1995). Time course of prefrontal lobe dysfunction in severely depressed in-patients: a longitudinal neuropsychological study. Psychological Medicine, 25, 79-85.*

Veiel, H. O. F. (1997). A preliminary profile of neuropsychological deficits associated with major depression. Journal of Clinical and Experimental Neuropsychology, 19, 587-603.

Wechsler, D. (1955). WAIS manual. New York: The Psychological Corporation.

Wertheimer, M. D. (1991). Affective disorders and organic mental disorders. International Psychogeriatrics, 3, 19-27.

Wolfe, J., Granholm, E., Butters, N., Saunders, E., & Janowsky, D. (1987). Verbal memory deficits associated with major affective disorders: a comparison of unipolar and bipolar patients. Journal of Affective Disorders, 13, 83-92.*

Zakzanis, K. K., Leach, L., & Kaplan, E. (1998). On the nature and pattern of neurocognitive function in major depressive disorder. Neuropsychiatry Neuropsychology and Behavioral Neurology, 11, 111-119.

Table 1.

Verbal Fluency Performance for Patients with Depression Relative to Healthy Controls

	M	K	N	SE	95% CIs; mean Lower	95% CIs; mean Upper	Z*	PVAF	Q*	σ_θ^2	SD	95% CIs; mean effects Lower	95% CIs; mean effects Upper
Phonemic fluency	**.30**	53	2179	.030	.24	.36	9.9	8.9	150.4	.029	.172	-.04	.64
Semantic fluency	**.44**	15	434	.052	.34	.54	8.4	19.2	29.1	.020	.143	.16	.72

*All values of Q and Z significant (ps < .05).

Table 2

Performance on Phonemic Fluency (PF), Semantic Fluency (SF) and other Cognitive Measures

	M	K	N	SE	95% CIs of mean Lower	95% CIs of mean Upper	Z	PVAF	Q	σ^2_θ	SD	95% CIs of mean effects Lower	95% CIs of mean effects Upper	
Studies that include PF														**PF M**
Semantic Fluency	**.43**	14	376	.056	.32	.54	7.7*	18.5	28.8*	.023	.15	.13	.73	**.39** (K = 14)
Premorbid IQ	**.19**	17	424	.081	.03	.35	2.4*	3.6	80.1*	.085	.29	-.38	.76	**.28** (K = 17)
VIQ	**.17**	15	727	.072	.03	.31	2.3*	2.8	72.3*	.059	.24	-.31	.64	**.26** (K = 15)
TMT A	**.24**	13	471	.049	.15	.34	4.9*	5.9	20.4	.013	.11	.02	.46	**.22** (K = 13)
Digit Symbol	**.21**	17	803	.052	.11	.32	4.2*	4.6	50.0*	.029	.17	-.12	.55	**.21** (K = 17)
BNT	**.05**	11	619	.058	-.07	.16	0.8	0.2	24.8*	.021	.14	-.23	.33	**.20** (K = 11)
Verbal Learning	**.35**	11	448	.060	.23	.46	5.8*	12.0	26.6*	.023	.15	.05	.65	**.19** (K = 11)
Delayed Recall	**.40**	14	480	.038	.33	.48	10.6*	16.3	16.2	.004	.06	.28	.53	**.28** (K = 14)
WCST CC	**.22**	14	744	.038	.14	.29	5.8*	4.7	18.3	.006	.07	.07	.36	**.26** (K = 14)
WCST PE	**.27**	14	744	.058	.16	.39	4.7*	7.5	47.4*	.032	.18	-.08	.63	**.26** (K = 14)
Stroop	**.17**	11	705	.050	.07	.26	3.3*	2.7	22.7*	.015	.12	-.08	.41	**.22** (K = 11)
Studies that include SF														**SF M**
Phonemic Fluency	**.39**	14	376	.054	.28	.50	7.3*	15.2	23.9*	.018	.13	.13	.65	**.43** (K = 14)
Premorbid IQ	**.27**	7	147	.159	-.04	.58	1.7	7.3	50.7*	.151	.39	-.49	1.00	**.46** (K = 7)
VIQ	**.42**	5	94	.133	.16	.69	3.2*	18.1	18.2*	.066	.26	-.08	.93	**.57** (K = 5)

*p < .05

Note; For each variable of interest, the mean effect for PF was recalculated using only those studies that assessed both; i.e. 17 studies assess both PF and premorbid intelligence; in addition to calculating a mean effect for premorbid IQ based on these 17 studies (r = .19), the mean effect for PF was also recalculated based only on these 17 studies (r = .28). Thus, when assessing whether or not the effect size for PF is in excess of that for premorbid intelligence, exactly the same participants have been tested upon each measure, effectively 'controlling' for any substantive differences between studies. This same procedure was adopted for comparisons of SF with other cognitive measures. This is a more rigorous method of comparing performance on different measures at the level of meta-analysis.

Table 3

Phonemic-Semantic Differential as a Function of Diagnostic Group

Diagnosis	**Phonemic r**	**Semantic r**	**Δ PVAF**
Depression	.39 (K = 14)	.43 (K = 14)	**3.28%**
DAT	.57 (K = 70)	.73 (K = 70)	**20.80%**

Appendix A: Studies Included in Quantitative Review

Study	Characteristics of Depressed Patients								Effect Size		Study Manipulation of Depressed Patients
	N	Criteria	Inpatient (%)	Medicated (%)	HD	Unipolar (%)	Psychotic (%)	Major (%)	Phonemic	Semantic	
Austin et al. (1999)	54	DSM	Majority	76	24.4[b]	-	24	100	.11		Melancholic
Austin et al. (1999)	23	DSM	Majority	57	20.2[b]	-	0	100	-.05		Non-melancholic
Austin et al. (1992)	20	DSM	-	55	24.1[a]	100	-	100	.27		Endogenous
Austin et al. (1992)	20	DSM	-	40	21.3[a]	100	-	100	.23		Neurotic
Basso et al. (2002)	25	DSM	100	Majority	-	0	32	-	.36		
Beatty et al. (1990)	14	DSM	0	0	-	-	0	100	.27	.13	
Blackwood et al. (1998)	10	-	100	-	-	100	-	100	.07		
Boone et al. (1994)	36	DSM	0	0	18.6[a]	-	0	100	.26		Age group 46-59
Boone et al. (1994)	23	DSM	0	0	21.4[a]	-	0	100	.25		Age group 60-69
Boone et al. (1994)	14	DSM	0	0	19.1[a]	-	0	100	.11		Age group 70-85
Boone et al. (1995)	37	DSM	0	0	15.6[a]	-	0	100	.12		Mild depression
Boone et al. (1995)	36	DSM	0	0	23.6[a]	-	0	100	.28		Moderate depression
Brown et al. (1994)	10	RDC	-	-	24.4[a]	Majority	-	100	.37	.42	Not impaired
Brown et al. (1994)	10	RDC	-	-	25.6[a]	Majority	-	100	.50	.43	Borderline
Brown et al. (1994)	9	RDC	-	-	25.1[a]	Majority	-	100	.74	.73	Impairment
Conaghan & Davidson (2002)	22	Other	100	-	-	-	0	-	.30		
Conaghan & Davidson (2002)	22	Other	-	-	-	-	0	100	.19		
Crews et al. (1999)	30	Other	0	0	-	-	-	100	-.11		
Crowe (1996)*	13	DSM	100	100	19.1[d]	-	-	46	.45		
Crowe & Hoogenraad (2000)	15	Various	20	-	-	-	-	20	.20		
de Groot et al. (1996)	25	DSM	100	0	29.1[a]	Majority	-	Majority	.67		Non-responders
de Groot et al. (1996)	26	DSM	100	0	25.5[a]	Majority	-	Majority	.61		Treatment responders
Degl'Innocenti et al. (1998)	17	DSM	100	0	-	-	-	100	.58		
Elliott et al. (1996)	28	DSM	39	100	22.4[a]	100	-	100	.41	.41	
Evangeli & Broks (2000)	12	Other	-	-	-	-	-	-	-.14		
Feinstein et al. (1998)	10	-	-	100	-	80	-	-	.09	-.13	
Fossati et al. (1999)	20	DSM	100	100	-	100	0	100	.21	.48	
Franke et al. (1993)	15	RDC	100	0	-	100	0	100	.38		
Gaudino et al. (1995)	18	**	0	0	-	-	-	-	.14		
Geffen et al. (1993)	10	DSM	100	0	-	100	-	100	.24	.55	
Hart et al. (1988)	17	DSM	-	0	31.0[c]	-	-	***	.47	.63	
Hart et al. (1987)	10	DSM	-	0	28.2[c]	-	-	***	.49	.54	
Ilsley et al. (1995)	7	DSM	-			-	0	100	.30	.34	Nonpsychotic
Ilsley et al. (1995)	8	DSM	-			-	100	100	.03	.15	Psychotic

Appendix A: Studies Included in Quantitative Review (continued)

Study	Characteristics of Depressed Patients								Effect Size		Study Manipulation of Depressed Patients
	N	Criteria	Inpatient (%)	Medicated (%)	HD	Unipolar (%)	Psychotic (%)	Major (%)	Phonemic	Semantic	
Johnson & Crockett (1982)	16	DSM	100	-	-	-	-	100	-.12		
King et al. (1991)	23	DSM	100	0	32.2[a]	100	-	100	.31	.33	
King et al. (1998)	57	DSM	100	67	28.9[c]	100	19	100	.23		
Kramer-Ginsberg et al. (1999)	23	DSM	-	-	-	100	-	100	.05		None/minimal WMH
Kramer-Ginsberg et al. (1999)	16	DSM	-	-	-	100	-	100	.52		More advanced WMH
Kuzis et al. (1997)	27	DSM	-	22	21.6[a]	-	-	100	.17		
Lafont et al. (1998)	16	DSM	-	75	-	-	-	100		.52	
Landro et al. (1997)	22	DSM	0	0	20.8[a]	100	0	100	.52		
Landro et al. (2001)	22	DSM	0	0	-	100	0	100	.32		
Lesser et al. (1991)	14	DSM	71	Majority	27.1[a]	93	100	100	.43		
Lyness et al. (1994)	18	DSM	0	0	26.9[a]	-	-	100	.08		Older
Lyness et al. (1994)	18	DSM	0	0	25.1[a]	-	-	100	.19		Middle-aged
Macleod & Salaminiou (2001)	22	Other	100	-	-	-	-	100	.28		
Macleod et al. (1997)	16	DSM	31	-	-	-	-	100	.30		
Nathan et al. (2001)	16	Other	-	-	-	-	-	-	.48		
Pendleton Jones et al. (1988)	10	DSM	-	90	30.5[a]	100	-	100	.16		
Shah et al. (1997)	14	DSM	100	73	23.5[a]	87	-	100	.28		
Smith et al. (1994)	26	DSM	0	-	-	-	0	100	.23		
Trichard et al. (1995)	23	DSM	100	78	-	-	-	-	.49	.53	
Wolfe et al. (1987)	12	DSM	100	0	23.8[a]	0	-	100	.47		

Hamilton depression rating scale total scores using 17-item scale[a], 21-item scale[b], or the 24-item scale[c], or the scale not specified[d].

* In Crowe's (1996) study, of the 13 patients, 6 had major depression, 6 dysthymic disorder and 1 schizoaffective disorder.

** In Gaudino et al.'s (1995) study, the patients are community volunteers who responded to an advertisement for individuals experiencing prolonged sadness; they were then found to be depressed on a self-rated measure of depression.

***Depressed patients meet criteria for DSM major affective disorder

NOTE: HD refers to Hamilton Depression Rating Scale; WMH refers to white matter hyperintensities