A Meta-Analytic Review of Verbal Fluency Deficits in Huntington’s Disease

Julie D. Henry, John R. Crawford, and Louise H. Phillips
King’s College, University of Aberdeen

A meta-analysis of 30 studies with 1,511 participants was conducted to estimate and compare the magnitude of deficits on tests of phonemic and semantic fluency for patients with Huntington’s disease (HD) relative to healthy control participants. As has been found for patients with focal frontal cortical lesions (but not for patients with focal temporal cortical lesions), symptomatic HD patients were comparably impaired on tests of phonemic and semantic fluency ($r_s = .71$ and $.73$, respectively). However, in contrast to patients with focal frontal lobe injuries, fluency deficits did not qualify as differential deficits relative to verbal intelligence or psychomotor speed. Therefore, for patients with HD, deficits on tests of phonemic and semantic fluency do not appear to reflect executive dysfunction but a more generalized cognitive impairment.

Keywords: executive, Huntington’s, meta-analysis

Although impairment in many cognitive domains has been documented in Huntington’s disease (HD; see Brandt, 1991), a very prominent view is that executive deficits characteristic of prefrontal dysfunction are a particularly striking feature of the disorder (Brandt, Bylsma, Aylward, Rothlind, & Gow, 1995; Caine, Hunt, Weingartner, & Ebert, 1978; Rosser & Hodges, 1994; Watkins et al., 2000). Consistent with this perspective are findings that HD patients often perform poorly on tests designed to capture executive dysfunction, such as the Wisconsin Card Sorting Test (WCST; Rebok, Bylsma, Keyl, Brandt, & Folstein, 1995), tests of verbal fluency (Rosser & Hodges, 1994), and the Stroop interference test (Hanes, Andrewes, Pantelis, & Chiu, 1996). Moreover, clinical observation and patient and caregiver reports have been interpreted as revealing a behavioral picture similar to that seen following frontal lesions (Caine et al., 1978), whereas, in neuro-pathological terms, HD is associated with subcortico-frontal abnormalities (Cummings, 1993). Because executive processes are presumed to rely heavily on the intact functions of frontal structures (e.g., see 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, it has been argued that although there is much anecdotal evidence that HD represents a dysexecutive syndrome, actual empirical support for this perspective is surprisingly sparse (Lawrence et al., 1998). Because fronto-executive hypotheses have been offered to account for the cognitive and behavioral disturbances seen in a vast array of neurological and psychiatric disorders in addition to HD, rigorous standards of proof must be applied when evaluating these hypotheses. Thus, because HD is characterized by multiple neurocognitive impairments, including deficits not associated with prefrontal dysfunction (Brandt, 1991), confirmation of a deficit on an executive task would, in isolation, provide only limited support for an executive hypothesis. Instead, it must be shown that the executive deficit is in excess of the average performance deficit across a range of other cognitive tasks that are not considered to impose heavy executive demands (Laws, 1999; Miller, 1984).

Verbal Fluency Deficits in HD

To address whether HD is particularly characterized by executive dysfunction, psychologists have extensively studied performance on tests of verbal fluency. For phonemic fluency, participants are asked to generate as many words as possible beginning with a specified letter (e.g., F); whereas for semantic fluency, search is constrained by a specified category (e.g., animals). Rosser and Hodges (1994) have argued that identical executive processes are involved in the initiation and monitoring of both of these tasks, and in a recent meta-analytic review Henry and Crawford (2004a) found quantitative support for this perspective. Although it is important to avoid conflating anatomy and cognition, as noted earlier, there is much evidence that executive processes are particularly dependent on frontal cortical structures. For example, Henry and Crawford (2004a) have found that focal frontal lobe injuries are associated with large but equivalent phonemic and semantic fluency deficits ($r_s = .52$ and $.54$, respectively), which suggests that the two types of tasks impose equivalent demands on executive control processes. However, Henry and Crawford (2004a) have found that, relative to patients with focal frontal lesions, semantic fluency was more impaired following focal temporal damage ($r = .61$), and this deficit was substantially larger than the corresponding phonemic fluency deficit ($r = .44$). Because there is much evidence supporting the notion that temporal structures are the neural substrates particularly responsible for semantic memory (see Fink & Randolph, 1998), this pattern of deficits was presumed to reflect the greater reliance of semantic fluency on the integrity of semantic memory.

The finding that performance on tests of phonemic and semantic fluency is often equivalently impaired in HD (see Monsch et al.,
Alzheimer’s Versus Huntington’s Dementias

Standard clinical dementia batteries are now capable of identifying distinct neurocognitive profiles that differentiate dementia of the Alzheimer type (DAT) from HD. However, it has been suggested that tests of phonemic and semantic fluency may also be of use in distinguishing between these two types of dementia (Rosser & Hodges, 1994), and this may be of particular clinical interest given their brevity and widespread use. A common assertion is that “cortical” dementias such as DAT are typified by a pattern of worse semantic, relative to phonemic, fluency performance, whereas “subcortical” dementias such as HD and PD are characterized by a pattern of comparable impairment on the two types of fluency (Hodges, Salmon, & Butters, 1990; Rosser & Hodges, 1994). However, not all researchers have found evidence consistent with this perspective. In a comparison of mildly demented patients with DAT, HD, and PD, for example, Suhr and Jones (1998) reported equivalent deficits on phonemic and semantic fluency in all three groups. In the present study, we compare the relative prominence of deficits on tests of phonemic and semantic fluency for patients with HD with the results found for studies involving DAT.

Cognitive Dysfunction in Preclinical HD

It also remains unclear whether there is preclinical cognitive dysfunction in HD. As Lawrence et al. (1998) point out, this issue is far from being resolved, with some studies reporting evidence of preclinical cognitive impairment (de Boo et al., 1997), and others failing to find any difference between mutation-positive and mutation-negative participants (Blackmore, Simpson, & Crawford, 1995). Identifying a possible prodromal phase in HD in which subtle but detectable behavioral and cognitive difficulties are present is of extreme importance to the gene carrier and may have therapeutic implications. In the present study, we explore the possibility that deficits in phonemic and semantic fluency are associated with preclinical HD by quantifying mean effects for each of these measures.

Aims of the Current Meta-Analysis

The first aim was to derive effect size estimates for phonemic and semantic fluency for patients with HD relative to healthy control participants. This is the first meta-analysis to compare the magnitude of deficits on these measures while ensuring that the same studies contribute to each, thereby permitting an extremely rigorous assessment of whether the verbal fluency deficit associated with HD predominantly reflects executive dysfunction or problems with semantic memory (Henry & Crawford, 2004a).

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; it may, instead, reflect a general verbal impairment or psychomotor slowing. Thus, the second aim was to estimate effect sizes for other cognitive measures in order to provide comparison standards and thus to assess the extent to which fluency deficits in HD qualify as differential deficits. We included premorbid intelligence as estimated by the National Adult Reading Test (NART; Nelson, 1982) and the reading subtest of the Wide Range Achievement Test (WRAT; Jastak & Wilkinson, 1984) to address the possibility that the
META-ANALYSIS OF FLUENCY IN HD

245

presence of a fluency deficit may reflect that patients with HD have not been successfully matched for premorbid ability with control participants. It is also important to address the possibility that fluency deficits simply reflect a current generalized verbal dysfunction (see Miller, 1984). Therefore, we compared the magnitude of the deficits on verbal fluency with the deficits on the VIQ and the FSIQ of the WAIS and WAIS–R (Wechsler, 1955, 1981) to determine whether the former qualifies as a differential deficit.

We also investigated whether deficits on tests of phonemic fluency are in excess of deficits on the WAIS Digit Symbol test (Wechsler, 1955, 1981), a widely used measure of psychomotor speed. We did this to address the possibility that deficits on tests of verbal fluency reflect the presence of bradyphrenia (i.e., a generalized reduction in cognitive speed) rather than executive dysfunction. We also compared performance on tests of phonemic and semantic fluency with the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983) because confrontation naming is considered to be very sensitive to the integrity of semantic memory (Hart, 1988), yet it imposes only minimal demands on effortful retrieval and cognitive speed.

For executive functioning, we recorded performance on Categories Completed (CC) and Perseverative Errors (PE) on the WCST (Heaton, 1981) and the interference condition of the Stroop (Golden, 1978). We hypothesized that a comparison of phonemic and semantic fluency with these other putative measures of executive function would be extremely informative both with respect to their convergent validity and their relative sensitivity to the presence of HD.

The third aim was to compare the relative prominence of phonemic and semantic fluency deficits in HD with the corresponding statistics for patients with DAT by comparing data from the present analyses with data that were taken from an independent meta-analysis (Henry, Crawford, & Phillips, 2004). Finally, the fourth aim was to quantify the mean effects for phonemic and semantic fluency for presymptomatic and symptomatic HD gene carriers.

Method

Sample of Studies

We performed a computer-based search involving the Web of Science, PsycLit [CD-ROM], and Science Direct databases 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. We also performed a manual search of most issues of the following journals published after 1969: Neuropsychology, Brain, Neuropsychologia, The Journal of the International Neuropsychological Society, The Clinical Neuropsychologist, Neuropsychiatry Neuropsychology and Behavioral Neurology, Journal of Neuropsychiatry and Clinical Neurosciences, and Journal of Clinical and Experimental Neuropsychology. It has been recommended that a manual search be conducted in addition to a computer-based search (Cooper, 1985; Green & Hall, 1984). As Green and Hall (1984) point out, despite its apparent cost in time, this is often a very effective manner of searching and may yield additional studies not found via other methods of search. This is particularly true in the present study because tests of verbal fluency are so routinely administered, often in conjunction with a large number of other neuropsychological tests, that they may not be referred to in an article’s title, abstract, or keywords. This means that a purely computer-based search would not identify many articles that are eligible for inclusion. The present search was completed in December 2002.

The inclusion criteria were first that the patient group had to consist entirely of adults with diagnoses of HD; all presymptomatic patients had to have undergone genetic testing and to have tested positive for the HD mutation. For the vast majority of studies that included symptomatic patients, neurologists made diagnoses on the basis of a positive family history, the presence of choreiform movements, and/or genetic testing. In addition, a study had to include a healthy control group free from neurological or psychiatric disease as well as a measure of phonemic and/or semantic fluency. Effect size estimates for premorbid IQ, current VIQ, current FSIQ Digit Symbol, BNT, WCST–CC, WCST–PE, and Stroop were derived from studies that also reported verbal fluency results. For inclusion, a study must have also presented precise statistics convertible to effect size r; thus, studies in which fluency data were derived from graphical information or imprecise p values were excluded (Butters, Sax, Montgomery, & Tarlow, 1978; Hodges et al., 1990; Randolph, Braun, Goldberg, & Chase, 1993). It should be noted that it was possible to include Snowden, Craufurd, Thompson, and Neary’s (2002) article in these analyses even though it did not provide precise statistics convertible to effect size r because Dr. Snowden provided us with the extra information we required. Finally, to be included, studies also had to have been published in English in a journal.

Statistical Analysis

Meta-analysis is a rigorous, quantitative alternative to the traditional review process because it involves statistical integration of results. The basis of this methodology is the effect size, a standardized statistic that quantifies the magnitude of an effect. In the present study, we used the effect size r, which corresponds to the degree of correlation between group membership (i.e., presence or absence of HD), and performance on the cognitive measure of interest. For each construct, we pooled effects to derive an estimate of the mean, with each effect weighted for sample size to correct for sampling error. To do so, we selected the random effects meta-analytic model (Shadish & Haddock, 1994) in preference to the more commonly used fixed effects model because it yields more generalizable 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 because it allows for substantive differences beyond sampling error that differentiate the effects contributing to each respective mean. The crucial statistical 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) has argued that the fixed effects model should be the exception rather than the rule because it may lead to inappropriately strong conclusions.

To estimate the degree of heterogeneity of the effects contributing to each mean, we estimated the homogeneity statistic Q and the random effects variance (r2), as well as the standard deviation of random effects and the 95% confidence intervals (CL) within which random effects can be expected to fall. Q quantifies within-group heterogeneity (i.e., the degree to which the effects contributing to each respective mean can be regarded as homogeneous). A significant Q statistic associated with a mean effect suggests that there are substantive differences between the effects contributing to that particular mean. In contrast, a nonsignificant estimate of Q suggests that once sampling error has been removed no substantive differences between the effects contributing to the respective mean in question remain (i.e., the null hypothesis of homogeneity of effects cannot be rejected).

It was also important to test whether the difference in the magnitude of mean effects between phonemic versus semantic fluency was statistically significant. However, there is no agreed upon method for statistically comparing mean effects by using the random effects meta-analytic model.
A particular difficulty is whether the degrees of freedom in such analyses should be based on the number of participants or the number of groups of participants (K). In the present work, t tests were computed with the more conservative K as the degree of freedom. It should be noted that we applied inferential statistics only for the phonemic–semantic comparison because for the other comparisons of interest K ranged from between two and nine and, thus, had low statistical power.

Because dementia severity moderates the magnitude of deficits across individual studies, for each comparison we included only studies that assessed both types of fluency. For example, although in total we assessed 27 independent groups of HD patients on phonemic fluency and 17 groups on semantic fluency, we only assessed 13 groups on both phonemic and semantic fluency; thus, when we compared phonemic and semantic fluency, we permitted only data from the latter 13 groups to contribute to the analyses. This effectively controls for the effect of illness severity (i.e., the same participants are compared on each of the measures of interest). We also calculated mean effects for each of the nonfluency variables identified (i.e., premorbid IQ, VIQ, FSIQ, Digit Symbol, BNT, WCST–CC, WCST–PE, and Stroop) and compared them with the corresponding effects for phonemic and semantic fluency. Again, to ensure that severity was controlled for, we included only studies that assessed both the fluency and nonfluency variable of interest in each comparison.

Finally, we tested the null hypothesis that the mean effect size is zero with the statistic \( Z \). Z values exceeding 1.96 indicate that the mean effect differs significantly from zero at the .05 level. To interpret how important a particular effect was in practical terms, we adopted Cohen’s (1977) guidelines. These suggest that a correlation of .10 should be regarded as a small effect size, .30 as a medium effect size, and .50 as a large effect size. In addition, we also presented squares of the effect size multiplied by 100 because these latter quantities represent the percentage of the variance accounted for (PVAF) on a measure of interest by group membership (i.e., the presence of HD vs. being a member of the healthy adult population).

**Results**

**Participant Characteristics**

Thirty research articles published between 1986 and 2002 contributed to the present study. In total, data from 727 HD patients and 784 control participants were included in these research articles. Supplementary information, including a complete list of the articles included in this meta-analysis, can be found at http://www.psy.unsw.edu.au/Scripts/AStaff.asp?Profile1De. Patients and control participants were closely matched for age (\( M = 43.6, SD = 7.28 \) vs. \( M = 43.7, SD = 8.72 \), respectively) and gender (45.4% male, 54.6% female vs. 44.9% male, 55.1% female, respectively). Although on average patients received only one half year less education than did control participants (\( M = 12.7 \) years, \( SD = 1.69 \) vs. \( M = 13.2 \) years, \( SD = 1.71 \), respectively), this difference attained significance, \( t(21) = 2.20, p = .039 \).

**Effect Sizes for Patients With HD Relative to Healthy Older Adults**

Table 1 presents estimates of the weighted mean effects for phonemic and semantic fluency, their variability, and the practical importance for symptomatic HD patients. Estimates of the weighted medians are also presented because these statistics are more robust to the presence of outliers than are mean values. For the mean and median effects, a positive sign indicates that patients have performed worse than did control participants. In the upper half of the table, mean effects are presented that have been calculated from any studies that included phonemic or semantic fluency (all studies). It can be seen that both these mean effects were significantly different from zero (\( p < .01 \)), and, in terms of practical importance as indexed by the PVAF, they were large in magnitude. However, for semantic fluency the mean effect (\( r = .73 \)) was larger than for phonemic fluency (\( r = .67 \)). For both mean effects, estimates of Q were highly significant (\( ps < .01 \)), reflecting the heterogeneous nature of the disorder. To avoid any potential confusion, it should be noted that these effect sizes differed from those presented in the present abstract and elsewhere because the latter effect sizes were based on studies that included both phonemic and semantic fluency, whereas the former effects were based on any studies that included phonemic and/or semantic fluency.

As Table 1 shows, when groups of participants for which both phonemic and semantic fluency measures were assessed were included in the analyses (\( K = 13 \)) the deficit for semantic fluency was only marginally larger than the deficit for phonemic fluency (\( rs = .73 \) vs. .71, respectively); this difference did not prove significant, \( t(12) = 0.31, p = .76 \).

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verbal Fluency Performance for Symptomatic Patients With HD Relative to Healthy Control Participants</strong></td>
</tr>
<tr>
<td><strong>Fluency</strong></td>
</tr>
<tr>
<td><strong>All studies</strong></td>
</tr>
<tr>
<td>Phonemic</td>
</tr>
<tr>
<td>Semantic</td>
</tr>
<tr>
<td><strong>Only studies that include phonemic and semantic fluency</strong></td>
</tr>
<tr>
<td>Phonemic</td>
</tr>
<tr>
<td>Semantic</td>
</tr>
</tbody>
</table>

*Note.* HD = Huntington’s disease; CI = confidence interval; PVAF = percentage of variance accounted for.

* \( p < .01 \)
Phonemic and Semantic Fluency Deficits Relative to Other Cognitive Deficits

Table 2 presents estimates of the mean effects, their variability, and the practical importance for premorbid IQ, current VIQ, FSIQ, Digit Symbol, BNT, WCST–CC, WCST–PE, and Stroop interference. We calculated these mean effects by using only those studies that included the particular measure of interest in addition to phonemic or semantic fluency. Thus, it can be seen that for each measure, the mean effects for phonemic fluency have been recalculated to ensure comparisons are fair (these effect sizes appear in the last column of Table 2). As noted previously, this methodology ensures that the same participants are contributing to the mean effects for the two variables being compared. This was particularly important because we found that the PVAF by phonemic and semantic fluency were both significantly and substantially negatively correlated with dementia severity as indexed by the Dementia Rating Scale (DRS; Mattis, 1988), indicating that as dementia severity increases deficits on the tests of fluency increase in magnitude (DRS with phonemic fluency: \( r = - .70, K = 10, p = .023 \); DRS with semantic fluency: \( r = - .85, K = 8, p = .007 \)).

Thus, Table 2 shows that for each nonfluency measure (e.g., premorbid intelligence), we calculated two mean effects: one for studies that also assess phonemic fluency (\( r = .19; K = 5 \)) and one for studies that also assess semantic fluency (\( r = - .02, K = 2 \)). Each fluency mean effect was also recalculated for these comparisons. All mean effects with the exception of those for premorbid intelligence were significant, and 6 of the 15 mean effects were associated with significant heterogeneity as indexed by \( Q (p < .05) \).

The effect sizes for both phonemic and semantic fluency were substantially in excess of the effect sizes for premorbid verbal intelligence (\( r_{s} = .51 \) and \( .19 \) for phonemic fluency vs. premorbid IQ; \( r_{s} = .57 \) and \( .02 \) for semantic fluency vs. premorbid IQ). However, there is no evidence that phonemic fluency is disproportionately impaired relative to either current VIQ or FSIQ (\( r_{s} = .65 \) and \( .63 \) for phonemic fluency vs. VIQ; \( r_{s} = .68 \) and \( .68 \) for phonemic fluency vs. FSIQ). Moreover, with respect to the measure of psychomotor speed (Digit Symbol), both the phonemic and semantic fluency deficits are of a smaller magnitude.

Finally, Table 2 also shows that phonemic fluency deficits were of a larger magnitude relative to the BNT, WCST–CC, and WCST–PE, but the Stroop interference was slightly more sensitive to the presence of HD. Semantic fluency deficits were larger relative to both the BNT and the Stroop interference test.

Phonemic and Semantic Fluency Deficits for Patients With HD Relative to DAT

Presented in Table 3 are the mean effects for phonemic and semantic fluency for patients with symptomatic HD and DAT (DAT data were taken from Henry et al., 2004). It can be seen that the semantic fluency deficit for patients with DAT is of an identical magnitude to the corresponding deficit for symptomatic patients with HD (\( r = .73 \)), whereas the phonemic fluency deficit is substantially smaller for DAT patients (\( r = .57 \) vs. .71, respectively).

Table 2

Performance on Tests of Verbal Fluency and Other Cognitive Measures for Symptomatic HD Patients Versus Healthy Control Participants

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>Mdn</th>
<th>K</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>Z</th>
<th>PVAF</th>
<th>Q</th>
<th>( \sigma_{Q}^{2} )</th>
<th>SD</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premorbid IQ</td>
<td>.19</td>
<td>.09</td>
<td>3</td>
<td>.149</td>
<td>−.10</td>
<td>.48</td>
<td>1.3</td>
<td>3.6</td>
<td>7.5</td>
<td>.048</td>
<td>.22</td>
<td>−.24</td>
<td>.62</td>
</tr>
<tr>
<td>VIQ</td>
<td>.63</td>
<td>.56</td>
<td>8</td>
<td>.063</td>
<td>.50</td>
<td>.75</td>
<td>10.0</td>
<td>39.3</td>
<td>30.1</td>
<td>.022</td>
<td>.15</td>
<td>.33</td>
<td>.92</td>
</tr>
<tr>
<td>FSIQ</td>
<td>.68</td>
<td>.68</td>
<td>4</td>
<td>.066</td>
<td>.55</td>
<td>.81</td>
<td>10.3</td>
<td>46.9</td>
<td>9.1</td>
<td>.010</td>
<td>.10</td>
<td>.49</td>
<td>.88</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>.84</td>
<td>.69</td>
<td>5</td>
<td>.051</td>
<td>.74</td>
<td>.94</td>
<td>16.5</td>
<td>70.6</td>
<td>7.9</td>
<td>.006</td>
<td>.08</td>
<td>.69</td>
<td>.99</td>
</tr>
<tr>
<td>BNT</td>
<td>.61</td>
<td>.62</td>
<td>4</td>
<td>.067</td>
<td>.48</td>
<td>.74</td>
<td>9.1</td>
<td>37.2</td>
<td>4.4</td>
<td>.006</td>
<td>.08</td>
<td>.46</td>
<td>.76</td>
</tr>
<tr>
<td>WCST-CC</td>
<td>.35</td>
<td>.46</td>
<td>5</td>
<td>.103</td>
<td>.15</td>
<td>.55</td>
<td>3.4</td>
<td>12.3</td>
<td>15.7</td>
<td>.038</td>
<td>.19</td>
<td>−.03</td>
<td>.73</td>
</tr>
<tr>
<td>WCST-PE</td>
<td>.45</td>
<td>.29</td>
<td>6</td>
<td>.105</td>
<td>.25</td>
<td>.66</td>
<td>4.3</td>
<td>20.3</td>
<td>25.4</td>
<td>.050</td>
<td>.22</td>
<td>.01</td>
<td>.89</td>
</tr>
<tr>
<td>Stroop interference</td>
<td>.61</td>
<td>.58</td>
<td>9</td>
<td>.052</td>
<td>.51</td>
<td>.72</td>
<td>11.8</td>
<td>37.7</td>
<td>16.7</td>
<td>.011</td>
<td>.11</td>
<td>.41</td>
<td>.82</td>
</tr>
</tbody>
</table>

Studies that include PF

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>Mdn</th>
<th>K</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>Z</th>
<th>PVAF</th>
<th>Q</th>
<th>( \sigma_{Q}^{2} )</th>
<th>SD</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premorbid IQ</td>
<td>−.02</td>
<td>.09</td>
<td>2</td>
<td>.143</td>
<td>−.30</td>
<td>.26</td>
<td>−.2</td>
<td>2.6</td>
<td>.026</td>
<td>.16</td>
<td>−.34</td>
<td>.29</td>
<td>.57</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>.89</td>
<td>.87</td>
<td>3</td>
<td>.049</td>
<td>.80</td>
<td>.99</td>
<td>18.4</td>
<td>80.0</td>
<td>2.4</td>
<td>.002</td>
<td>.04</td>
<td>.82</td>
<td>.97</td>
</tr>
<tr>
<td>BNT</td>
<td>.61</td>
<td>.62</td>
<td>5</td>
<td>.050</td>
<td>.52</td>
<td>.71</td>
<td>12.4</td>
<td>37.6</td>
<td>4.5</td>
<td>.001</td>
<td>.04</td>
<td>.54</td>
<td>.69</td>
</tr>
<tr>
<td>Stroop interference</td>
<td>.52</td>
<td>.41</td>
<td>5</td>
<td>.072</td>
<td>.38</td>
<td>.66</td>
<td>7.2</td>
<td>27.2</td>
<td>5.6</td>
<td>.008</td>
<td>.09</td>
<td>.35</td>
<td>.69</td>
</tr>
</tbody>
</table>

Studies that include SF

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>Mdn</th>
<th>K</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>Z</th>
<th>PVAF</th>
<th>Q</th>
<th>( \sigma_{Q}^{2} )</th>
<th>SD</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premorbid IQ</td>
<td>−.02</td>
<td>.09</td>
<td>2</td>
<td>.143</td>
<td>−.30</td>
<td>.26</td>
<td>−.2</td>
<td>2.6</td>
<td>.026</td>
<td>.16</td>
<td>−.34</td>
<td>.29</td>
<td>.57</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>.89</td>
<td>.87</td>
<td>3</td>
<td>.049</td>
<td>.80</td>
<td>.99</td>
<td>18.4</td>
<td>80.0</td>
<td>2.4</td>
<td>.002</td>
<td>.04</td>
<td>.82</td>
<td>.97</td>
</tr>
<tr>
<td>BNT</td>
<td>.61</td>
<td>.62</td>
<td>5</td>
<td>.050</td>
<td>.52</td>
<td>.71</td>
<td>12.4</td>
<td>37.6</td>
<td>4.5</td>
<td>.001</td>
<td>.04</td>
<td>.54</td>
<td>.69</td>
</tr>
<tr>
<td>Stroop interference</td>
<td>.52</td>
<td>.41</td>
<td>5</td>
<td>.072</td>
<td>.38</td>
<td>.66</td>
<td>7.2</td>
<td>27.2</td>
<td>5.6</td>
<td>.008</td>
<td>.09</td>
<td>.35</td>
<td>.69</td>
</tr>
</tbody>
</table>

Note. The mean effects for phonemic and semantic fluency were recalculated for each comparison of interest. For example, only nine studies included both phonemic fluency and Stroop interference. In addition to calculating the mean effect for Stroop interference from these nine studies (\( r = .61 \)), the mean effect for phonemic fluency was also recalculated on the basis of only these nine studies (i.e., \( r = .57 \)). Thus, in each comparison exactly the same participants have been tested on each of the measures of interest, effectively controlling for any substantive differences between studies, such as the stage of the disease. HD = Huntington’s disease; CI = confidence interval; PVAF = percentage of variance accounted for; PF = phonemic fluency; VIQ = verbal IQ; FSIQ = full scale IQ; BNT = Boston Naming Test; WCST–CC = Wisconsin Card Sorting Test—Categories Computed; WCST–PE = Wisconsin Card Sorting Test—Perseverative Errors; SF = semantic fluency.

* \( p < .05 \).
In Table 3, mean effects for phonemic and semantic fluency are also presented for presymptomatic patients with HD. It can be seen that there is evidence of a slight deficit for measures of phonemic and semantic fluency even in the preclinical stage of the disorder, but these are small in magnitude (r1.12 vs. r1.17, respectively). However, in terms of the PVAF, the semantic fluency deficit is over twice as large as the phonemic fluency deficit for patients in the preclinical stage of the disease.

It is of interest that for symptomatic HD patients the difference in terms of the PVAF between phonemic and semantic fluency is substantially (although not significantly) positively correlated with dementia severity as indexed by the DRS (r=.62, K=7, p=.140). This suggests that the higher the score on the DRS (i.e., the less demented the patient), the larger the difference between phonemic and semantic fluency.

Assessing the Possibility of Publication Bias

A number of validity threats have been identified that may lead to imprecise conclusions in both nonquantitative 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 are nonsignificant results (Easterbrook, Berlin, Gopalan, & Mathews, 1991). To assess whether this bias posed a threat to the results of the present study, funnel plot diagrams were constructed for each of the fluency and nonfluency measures of interest. In these diagrams, sample size was plotted against the corresponding study-level effect; if statistically nonsignificant 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.

Table 3
Mean Fluency Effect Sizes for Symptomatic and Presymptomatic HD Patients and Patients With DAT

<table>
<thead>
<tr>
<th>Fluency</th>
<th>M</th>
<th>Mdn</th>
<th>K</th>
<th>N</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>Z</th>
<th>PVAF</th>
<th>Q</th>
<th>2SD</th>
<th>95% CIs of mean</th>
<th>95% CIs of mean effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symptomatic HD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic</td>
<td>.71</td>
<td>.73</td>
<td>13</td>
<td>443</td>
<td>.043</td>
<td>.62</td>
<td>.79</td>
<td>16.5*</td>
<td>50.0</td>
<td>48.0*</td>
<td>.015</td>
<td>.12</td>
<td>.47</td>
</tr>
<tr>
<td>Semantic</td>
<td>.73</td>
<td>.74</td>
<td>13</td>
<td>443</td>
<td>.027</td>
<td>.68</td>
<td>.79</td>
<td>26.8*</td>
<td>53.9</td>
<td>19.5</td>
<td>.003</td>
<td>.06</td>
<td>.62</td>
</tr>
<tr>
<td>Presymptomatic HD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic</td>
<td>.12</td>
<td>.12</td>
<td>4</td>
<td>208</td>
<td>.069</td>
<td>−.02</td>
<td>.25</td>
<td>1.7</td>
<td>1.3</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>.17</td>
<td>.14</td>
<td>4</td>
<td>208</td>
<td>.067</td>
<td>.04</td>
<td>.30</td>
<td>2.5*</td>
<td>2.9</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic</td>
<td>.57</td>
<td>.52</td>
<td>70</td>
<td>2,674</td>
<td>.024</td>
<td>.52</td>
<td>.62</td>
<td>23.8*</td>
<td>32.6</td>
<td>600.3*</td>
<td>.033</td>
<td>.180</td>
<td>.22</td>
</tr>
<tr>
<td>Semantic</td>
<td>.73</td>
<td>.68</td>
<td>70</td>
<td>2,674</td>
<td>.017</td>
<td>.69</td>
<td>.76</td>
<td>42.7*</td>
<td>52.7</td>
<td>630.6*</td>
<td>.016</td>
<td>.128</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note. Blank cells indicate that the random effects variance component is estimated to be zero. HD = Huntington’s disease; DAT = dementia of the Alzheimer type; PVAF = percentage of variance accounted for.

*p < .05.

Phonemic and Semantic Fluency Deficits for
Presymptomatic HD Patients

In Table 3, mean effects for phonemic and semantic fluency are also presented for presymptomatic patients with HD. It can be seen that there is evidence of a slight deficit for measures of phonemic and semantic fluency even in the preclinical stage of the disorder, but these are small in magnitude (rs=.12 vs. .17, respectively). However, in terms of the PVAF, the semantic fluency deficit is over twice as large as the phonemic fluency deficit for patients in the preclinical stage of the disease.

It is of interest that for symptomatic HD patients the difference in terms of the PVAF between phonemic and semantic fluency is substantially (although not significantly) positively correlated with dementia severity as indexed by the DRS (r=.62, K=7, p=.140). This suggests that the higher the score on the DRS (i.e., the less demented the patient), the larger the difference between phonemic and semantic fluency.

Discussion

Verbal Fluency Deficits in HD

A prominent view in the literature is that HD should be regarded as a disorder particularly characterized by fronto-executive dysfunction (Caine et al., 1978; Watkins et al., 2000) and the pattern of comparable impairment on tests of phonemic and semantic fluency that is often reported has been regarded as evidence consistent with this possibility (Rosser & Hodges, 1994). However, when mean effects were calculated for phonemic and semantic fluency including all studies that assessed either of these measures, it was found that the semantic fluency deficit was the notably larger of the two (rs=.73 vs. .67, respectively), as was found in an earlier meta-analytic review by Zakzanis (1998).

However, as noted earlier, if comparisons are to be fair 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. This rules out, for instance, the possibility that the groups of patients contributing to the mean for the former are more severely demented than those contributing to the mean for the latter. When we recalculated mean effects in the present study using a more rigorous methodology based only on studies that had assessed both phonemic and semantic fluency, we found that there was virtually no difference in the relative sensitivity of the two measures to the presence of HD (rs=.71 and .73 for phonemic and semantic fluency, respectively).

Evidence for a Differential Executive Deficit in HD?

As discussed previously, a pattern of comparable impairment on tests of phonemic and semantic fluency for patients with HD may reflect executive dysfunction if 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 on premorbid verbal IQ, they were virtually identical in magnitude to those observed for current VIQ and FSIQ. Both phonemic and semantic fluency were also substantially less impaired than was Digit Symbol, a measure of psychomotor speed. It is of interest that in a meta-analytic review of verbal fluency deficits in PD it was also found that, relative to Digit Symbol, neither phonemic nor semantic fluency were disproportionately impaired (Henry & Crawford, 2004c). As in that study, conclusions must be cautiously drawn given that very few studies contributed to the analyses involving Digit Symbol; however, these results are at least consistent with the possibility that for patients with subcortical dementias such as HD and PD, bradyphrenia or cognitive slowing may partially underlie fluency deficits.

There is much evidence that VIQ is a strong predictor of fluency performance and that fluency tasks impose substantial demands on both language abilities and psychomotor speed (see Salthouse, Atkinson, & Berish, 2003). Thus, for some patient groups fluency deficits may not primarily reflect executive dysfunction but instead verbal or motor speed problems. Because fluency tests (like all cognitive measures) are multifactorial, deficits across different patient groups may not reflect the same underlying cognitive impairment (see Miller, 1984). Thus, we argue that it is the relative magnitude of deficits that is important. When deficits on tests of fluency exceed deficits on verbal ability and psychomotor speed, as we have found for both patients with focal frontal cortical lesions as well as patients with TBI (Henry & Crawford, 2004a, 2004b), it is consistent with the possibility that the cognitive impairment underlying the fluency deficit may be executive dysfunction.

Given that HD is characterized by multiple neurocognitive impairments, deficits of verbal intelligence and motor speed would be expected. The evidence presented here for patients with HD shows that the fluency deficits are not differentially larger in magnitude than the deficits in verbal intelligence or speed, suggesting that fluency deficits in HD may stem not primarily from executive dysfunction but slowed cognitive processing speed and/or a deficit in verbal ability. Thus, although patients with focal frontal injuries and HD may be similar in that they exhibit equivalent deficits across measures of phonemic and semantic fluency, the impairment underlying these fluency deficits appears to differ across these patient groups. We suggest that only for the former group do fluency deficits primarily reflect executive dysfunction.

However, it should be noted that VIQ as measured by the WAIS may also impose some limited demands on executive processes. It is often argued that the WAIS is relatively insensitive to the effects of prefrontal dysfunction (Dempster, 1992; Lezak, 1995; Stuss & Benson, 1986). However, others have argued that the insensitivity of the WAIS to prefrontal dysfunction may have been exaggerated (Parker & Crawford, 1992; Shallice, 1988). Nevertheless, the present rationale does not require that the WAIS be entirely insensitive; we simply suggest that it is less sensitive than validated executive tasks.

It could also be argued that a differential deficit in executive functioning does exist in HD but that the fluency tests were not sensitive enough to expose this; that is, other executive measures may reveal such a deficit. Two facts argue against this alternative interpretation: A very substantial differential deficit on fluency relative to IQ is observed following focal frontal lesions (Henry & Crawford, 2004a) and TBI (Henry & Crawford, 2004b). Furthermore, in the present study, phonemic fluency was much more sensitive to the presence of HD than the WCST, a widely used alternative measure of executive functioning.

**Deficits as a Function of Semantic Fluency Category Type**

It has been suggested that the retrieval demands of semantic fluency tasks may vary according to the semantic category involved, and in particular, a distinction has been made between broad semantic categories such as *supermarket items* and more constrained semantic categories such as *fruits and vegetables* (Randolph et al., 1993), with the latter thought to impose lower demands on effortful retrieval processes. Thus, the neurocognitive demands of some semantic fluency tasks may be more or less similar to phonemic fluency tasks, depending on the categories involved. Although it would be of considerable interest to quantify the effect sizes for each of the different types of categories, this in practice is not possible from the information provided in the primary studies. For the vast majority of studies that measure semantic fluency, the “animal fluency” category has been used either on its own or as part of a larger group of categories. None of the studies that assessed more than one type of semantic fluency reported scores for each of the categories individually. Exploring whether the type of phonemic fluency that is used moderates the magnitude of the deficit is also not possible given that the vast majority of studies use the variant FAS. However, the possibility that the variant of phonemic or semantic fluency used moderates the magnitude of the deficit in HD should be investigated in future primary research.

**Sources of Heterogeneity in HD**

For many of the variables assessed in the present study, the homogeneity statistic $Q$ was significant. The fact that corrections have been implemented for sampling error, (the most serious source of artificial variance; Hunter, Schmidt, & Jackson, 1982) suggests that substantive differences between studies remain. For the majority of these mean effects, the significant heterogeneity observed cannot be attributed to the presence of a few extreme values as only two outliers were identified in the present study. Because exclusion of these outliers did not alter the basic pattern of results observed we retained them.

Nevertheless, it is clear that there is considerable heterogeneity in the effects contributing to many of the mean effects. The magnitude of the heterogeneity observed for many of the mean effects reinforces the importance of adopting a conservative meta-analytic methodology of the type used in the present study. As noted earlier, if two measures are to be compared fairly, it is important that the patients contributing to the mean effect for one do not differ from the patients contributing to the other. In the present meta-analysis, a very obvious source of heterogeneity across studies is likely to be dementia status, as the patients contributing to the present meta-analysis varied markedly in this regard. As noted, both phonemic and semantic fluency were very significantly and substantially correlated with dementia severity as indexed by the DRS ($r = -.70$ and $-.85$, respectively).

However, it is likely that in addition to dementia severity there are other important moderators of these effects. It has been sug-
gested, for instance, that age of onset (Gomez-Tortosa et al., 1998) may moderate the magnitude of cognitive deficits observed. Unfortunately, it was not possible to explore the influence of other such potential moderators in the present study as very few studies presented the prerequisite data. However, the degree of heterogeneity identified in the present study very strongly points to the importance of exploring the influence of such variables in future primary research. Thus, although in general, patients with HD should be comparably impaired on semantic and phonemic fluency, this may not be the case for certain subgroups.

Of course, it is also possible that the large Q values observed at least partially reflect heterogeneity across the control participants sampled. It may, for instance, reflect differences in matching controls to HD participants across demographic variables such as age, education, or gender. A review of the control participant profiles, however, suggests that this is unlikely to be the case. In virtually all the studies HD patients and controls were matched across most, if not all, of these demographic variables. On the few occasions in which they were not, or when this information was not provided, there was no evidence that the magnitude of the effect sizes for phonemic or semantic fluency varied systematically from studies that adopted more rigorous matching criteria.

**Distinguishing DAT and HD**

Although patients with HD and DAT performed equally poorly on a measure of semantic fluency (for both groups, \( r = .73 \)), the latter group was substantially less impaired on phonemic fluency (\( r_s = .57 \) vs. \( .71 \), respectively). Thus, patients with HD exhibit comparable impairment on the two types of fluency, whereas patients with DAT are substantially more impaired on semantic relative to phonemic fluency. Indeed, the difference in terms of the PVAF between phonemic and semantic fluency was estimated to be 20.8% for patients with DAT but was estimated to be only 2.9% for patients with HD. It is of interest that a similar pattern of results was found when comparing the results of meta-analytic reviews involving patients with PD and DAT (Henry & Crawford, 2004c). Thus, the results across both of these earlier, as well as the present, meta-analyses do suggest that differentiating between cortical dementias such as DAT and subcortical dementias such as PD and HD may be possible by assessing the relative prominence of deficits on tests of phonemic and semantic fluency.

However, Brown and Marsden (1998) have argued that many of the suggested differences between cortical and subcortical dementias can be attributed to differences in dementia severity, and they have questioned the idea that there are distinctive patterns of cognitive impairment for the two types of dementia. This is a difficult issue to resolve in the present meta-analysis because the extent to which the dementia severity of the HD patients contributing to the HD analyses compares with the dementia severity of the DAT patients contributing to the DAT analyses is unclear. Many of the studies have not reported scores on a standardized measure of dementia, and when studies have reported these statistics the same measures are not often used. For example, in the HD analyses the DRS has been the most commonly reported measure of dementia severity, whereas in the DAT analyses the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975) has been typically used for this purpose.

Nevertheless, it is quite striking that the semantic fluency deficit for patients with both HD and DAT was identical in magnitude (\( r = .73 \)). This suggests that the larger deficit on phonemic fluency for the HD relative to the DAT group was not simply because the former patient group are more severely demented. Moreover, it is interesting to note that in a sample of patients carefully matched for overall level of dementia, precisely this pattern of results has been found (Rosser & Hodges, 1994).

However, it is of interest that for patients with DAT the difference in terms of the relative sensitivity of phonemic and semantic fluency to the presence of HD as indexed by the PVAF was not significantly or substantially correlated with dementia severity (\( r = .16 \), \( K = 45 \), \( p = .29 \)), whereas for patients with HD, the difference in terms of the PVAF was substantially (although not significantly) correlated with dementia severity on the DRS (\( r = .62 \), \( K = 7 \), \( p = .14 \)). Although speculative, particularly because the number of studies contributing to the latter statistic was small in number, it may be that the overlap between HD and DAT in terms of the relative prominence of deficits on phonemic and semantic fluency is greater in the earlier stages of the disease. That is, for patients with HD in the early stages of dementia there may be a tendency for semantic fluency to be more impaired than phonemic fluency, but as the dementia progresses these deficits become more comparable in magnitude. However, for patients with DAT, semantic fluency is more impaired than phonemic fluency at every stage of the disorder to a comparable degree. Indeed, it is of interest that in their comparison of patients with DAT, HD, and PD, Suhr and Jones (1998) found that for all three groups parallel semantic/letter fluency deficits were found for patients who were only mildly demented. The results of the present meta-analysis suggest that the use of the relative prominence of phonemic and semantic fluency deficits to distinguish between HD and DAT may be more difficult in the earlier stages of the disorders but should clearly differentiate them as the dementia grows more advanced.

**Cognitive Impairment in Preclinical HD**

As noted previously, earlier empirical studies have failed to agree on whether presymptomatic HD is associated with cognitive dysfunction (Blackmore et al., 1995; Lawrence et al., 1998). The present results are consistent with Blackmore et al.’s (1995) argument that minimal deficits may be apparent in preclinical HD, but existing measures of assessment are not sensitive enough to identify them. Although there is evidence of a verbal fluency deficit in preclinical HD, the magnitude of this deficit is so small that many studies may simply lack the power to detect a statistically significant effect.

However, consistent with the possibility that semantic fluency may be more impaired than phonemic fluency in the earlier stages of HD, patients in the preclinical stages of the disorder were found to be substantially more impaired on the former measure. Although the absolute magnitude of the deficits on both these measures were small (\( r_s = .12 \) and \( .17 \) for phonemic and semantic fluency, respectively), in terms of the PVAF the latter measure was over twice as large (1.3% vs. 2.9%). Thus, preclinical HD does appear to be associated with cognitive dysfunction, but this may differ qualitatively as well as quantitatively from patients in the ad-
advanced stages of HD because the latter group typically present with comparable deficits on both types of fluency.

Limitations of the Present Study

The results of the present study contradict the prevailing view in the literature and suggest that HD is not characterized by disproportionate executive dysfunction. However, it is important to emphasize that the estimates of $Q$ for many of the mean effects indicate that there is considerable heterogeneity across the studies contributing to the present meta-analysis. Moreover, many of the mean effects have been calculated from a small number of studies. As Rosenthal and DiMatteo (2002) note, mean effects derived from a small number of studies must be regarded as relatively unreliable. Thus although the present results provide evidence consistent with the possibility that fluency deficits in HD do not qualify as differential deficits relative to deficits on other cognitive measures such as VIQ, further research is needed to cross-validate and test the generalizability of these findings.

Summary and Conclusions

As has been found for frontal patients and patients who have sustained a TBI, HD patients were comparably impaired on tests of phonemic and semantic fluency. However, in contrast to these other patient groups, for patients with HD fluency deficits did not qualify as differential deficits relative to measures that do not heavily load executive functions (e.g., verbal intelligence and psychomotor speed). Thus, in contrast to the prevailing view in the literature, there is no evidence that HD is particularly characterized by executive dysfunction, at least as indexed by tests of phonemic and semantic fluency. Relative to patients with DAT, patients with HD were comparably impaired on a measure of semantic fluency, but the former group was substantially less impaired on phonemic fluency, indicating that these two etiologically distinct types of dementia may be differentiated from one another by the relative prominence of deficits on these two measures. For patients in the preclinical stage of HD, although only small in magnitude, deficits were found on tests of phonemic and semantic fluency.

References


Received January 16, 2004
Revision received March 30, 2004
Accepted April 14, 2004