Measuring Executive Dysfunction Longitudinally and in Relation to Genetic Burden, Brain Volumetrics, and Depression in Prodromal Huntington Disease

Kathryn V. Papp, Peter J. Snyder, James A. Mills, Kevin Duff, Holly J. Westervelt, Jeffrey D. Long, Spencer Lourens, Jane S. Paulsen

1Department of Psychiatry, Harvard Medical School, Boston, MA 02215, USA
2Department of Neurology, Alpert Medical School of Brown University & Lifespan Hospitals System, Providence, RI 02903, USA
3Department of Psychiatry, University of Iowa, Iowa City, IA 52242, USA
4Department of Neurology, University of Utah, Salt Lake City, UT 84108, USA
5Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University & Department of Psychiatry, Rhode Island Hospital, Providence, RI 02903, USA
6Department of Biostatistics, University of Iowa, Iowa City, IA 52242, USA

*Corresponding author at: The University of Iowa, Roy J. and Lucille A. Carver College of Medicine, Research, 1-305 Medical Education Building, Iowa City, IA 52242-1000, USA. Tel.: 319-353-4551; fax: 319-353-3003. E-mail address: predict-publications@uiowa.edu (J. S. Paulsen).

Accepted 31 October 2012

Abstract

Executive dysfunction (ED) is a characteristic of Huntington disease (HD), but its severity and progression is less understood in the prodromal phase, e.g., before gross motor abnormalities. We examined planning and problem-solving abilities using the Towers Task in HD mutation-positive individuals without motor symptoms (n = 781) and controls (n = 212). Participants with greater disease progression (determined using mutation size and current age) performed more slowly and with less accuracy on the Towers Task. Performance accuracy was negatively related to striatal volume while both accuracy and working memory were negatively related to frontal white matter volume. Disease progression at baseline was not associated with longitudinal performance over 4 years. Whereas the baseline findings indicate that ED becomes more prevalent with greater disease progression in prodromal HD and can be quantified using the Towers task, the absence of notable longitudinal findings indicates that the Towers Task exhibits limited sensitivity to cognitive decline in this population.

Keywords: Huntington’s disease; Genetic disorders; Executive functions; Neuroimaging (structural); Norms/normative studies; Practice effects/reliable change; longitudinal change

Introduction

Huntington disease (HD) is a progressive autosomal-dominant neurodegenerative disorder characterized by a triad of symptoms, including motor dysfunction, psychiatric disturbance, and cognitive deficits. The abnormality results from the expansion of the normal polyglutamine cytosine–adenine–guanine (CAG) repeat (36 or more repeats) within a gene on the short arm of chromosome 4 that codes for the protein huntingtin (The Huntington’s Disease Collaborative Research Group, 1993). The principal site of neuropathology in HD is the caudate nucleus that receives its primary afferent projections from the dorsolateral prefrontal cortex. The preservation of these pathways is associated with a specific area of executive functioning: problem-solving requiring strategy development and working memory (Bechara, Damasio, Tranel, & Anderson, 1998; Sullivan, Riccio, & Castillo, 2009 for a review). As expected, those diagnosed with HD exhibit difficulty with tests that tap into these types of executive functions (Hanes, Andrewes, Smith, & Pantelis, 1996; Paulsen et al., 1996; Watkins et al., 2000).
The Towers Task is a measure of problem-solving, strategy development, and working memory. It has evolved into a number of modified versions with unique norms, instructions, and scoring criteria from the original Tower of Hanoi (Simon, 1975) and Tower of London (Shallice, 1982) tasks. All versions involve rearranging balls, beads, or disks to achieve a pre-specified end result using the fewest number of moves possible. Depending on the version used, multiple metrics of performance (accuracy, time, rule violations) are available. A recent meta-analysis indicates that the measures of accuracy most able to distinguish between neuropsychological populations and controls include (i) number of trials completed and (ii) number of legal moves (Sullivan et al., 2009). The most sensitive measure for timing variables was Execution Time (time between first move and task completion) followed by Initiation Time (time to first move, commonly considered “planning”) and Total Time. The task requires planning and use of a strategy: if a participant visualizes the next group of steps before executing them, he is able to make fewer errors and therefore use fewer moves (Levin et al., 1991; Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997). The Towers Task involves a working memory component because a sequence of moves must be stored and used during the task. Research has indicated that as the Tower becomes more difficult (more pegs or beads), working memory deficits become associated with decreased performance (Goel, Pullara, & Grafman, 2001).

The goal of our study was to examine whether the Towers Task could detect executive dysfunction (ED) in gene-expanded individuals who do not yet meet criteria for a clinical diagnosis of HD at time of study entry, for example, prodromal HD. We also sought to examine the contribution of different aspects of executive functions (i.e., planning, working memory, accuracy) to performance in this population. The detection of ED in prodromal HD has been variable (Georgiou-Karistianis et al., 2003; Hahn-Barma et al., 1998; Ho et al., 2003; Lemiere, Decruyenaere, Evers-Kiebooms, Vandenbussche, & Dom, 2004). This variability is likely the result of several methodological and measurement limitations, including (i) the use of various executive function measures with various levels of sensitivity, (ii) attempts to compare both cross-sectional and longitudinal designs, (iii) difficulty controlling for practice or “spoiling” effects on executive measures in longitudinal designs, (iv) difficulty in parsing out the secondary effects of motor speed, and (v) insufficiently powered sample sizes, (vi) varying samples of prodromal and diagnosed HD samples used without reliable grouping criteria, and (vii) perhaps most importantly, the fact that executive functions are a broad label for a diverse set of functions. The current study attempted to address a number of these prior limitations.

We chose to look at performance on the Towers Task in a large sample in order to determine the sensitivity of the task to different stages in the HD prodrome. We also used a computerized measure, which benefits from extremely reliable administration and millisecond timing accuracy (Schatz & Browndyke, 2002). We accounted for the potentially confounding effect of motor speed performance in this population by co-varying with a pure motor task. We were additionally able to examine different aspects of executive functions by examining Towers outcomes representing different cognitive skills (accuracy vs. efficiency) and to look at the effect of working memory load by adjusting the difficulty of the task.

Given the sensitivity of the Towers Task in diagnosed HD (Cohen’s d effect size of 1.5 comparing HD with controls in Hanes et al., 1996; correlation \( r = .647 \) between Tower of Hanoi and reduced caudate volume in Peinemann et al., 2005), it is important to determine whether this measure is also sensitive clinically, cross-sectionally, and longitudinally in the HD prodrome. We hypothesize that those with greater disease progression at time of study entry (i.e., those closer to estimated clinical motor diagnosis and those showing reduced striatal and frontal white matter volumes) would perform more slowly and less accurately than those who were less progressed. We also expect to see declines in performance over time given that HD is a neurodegenerative process.

This study examined performance on the Towers Task with three and four disk problems that allowed the relationship between performance and cognitive demand (Shallice, 1982; Goel et al., 2001) to be analyzed. Previous research indicates that changes in ED in prodromal HD are subtle (Papp, Kaplan, & Snyder, 2011). We therefore hypothesize that Towers 4 (four disks) would be more sensitive to detecting subtle differences between groups. Previous research indicates a relationship between the number of disks used and working memory load (Goel et al., 2001), and thus, Towers 4 has a higher difficulty level compared with Towers 3. While cross-sectional differences in performance have previously been reported for one Towers outcome using this study sample (Stout et al., 2011), here we examine multiple outcome metrics of Towers performance, including timing and accuracy (moves required), which provides the opportunity to better understand the components of Towers performance and how it varies during the HD prodrome. Furthermore, we compared Towers performance across stages of disease progression in the HD prodrome and with brain volumetrics (striatum and frontal white matter). Previous research has shown that gene-expanded individuals within 15 years of predicted diagnosis show reduced total brain tissue, cerebral spinal fluid, white matter, cortical gray matter, thalamus, caudate, and putamen compared with non-affected relatives. Total striatal volume demonstrated the largest differences between non-affected family members and prodromal individuals followed by cerebral white matter (Paulsen et al. 2010). While we focus on a performance-based approach to measure executive functions, we also examine the ecological and convergent validity of direct measurement of cognition by examining the relationship between cognitive performance and self- and companion reports of ED. Previous research has shown that self- and companion report of “frontal” behaviors has been associated with proximity to HD diagnosis (Duff et al., 2010). We expect Towers
performance to be correlated with questionnaire measures of ED completed by companions, but not by participants. This is based on previous findings suggesting an increased lack of insight for cognitive difficulties in those with fronto-subcortical disruption, which are evident even in prodromal HD (Duff et al., 2010). Given the known effects of depression on cognition (Nehl, Ready, Hamilton & Paulsen, 2001), Towers performance was also examined with regard to measures of mood symptomatology.

**Methods**

**Participants**

Information about inclusion/exclusion criteria was obtained through a combination of clinical interviews and self-report questionnaires. Participants included two groups of individuals from affected families: (i) those with the gene expansion (CAG ≥36) but without motor signs sufficient for a clinical HD diagnosis at study entry and (ii) those without the gene expansion (CAG <36). Participants were excluded if they had a positive history of other CNS disease or events (e.g., seizures or head trauma), pacemaker, metallic implants, prescribed antipsychotic or phenothiazine-derivative antiemetic medication in the past 6 months, and clinical evidence of unstable medical or psychiatric illness. Additionally, individuals with significant developmental cognitive disorders (e.g., mental retardation, special education for reading or math) were excluded. No restrictions were imposed regarding over-the-counter and natural remedies. All participants were 18 years or older and had voluntary independent genetic testing prior to enrollment in the study. Confirmation of the CAG repeat length was determined from baseline blood draws.

The current analysis examined the baseline and follow-up visits over 5 years of up to 781 gene-expanded individuals and 212 gene non-expanded individuals who enrolled between October 2002 and October 2009. Gene-expanded individuals were grouped according to their baseline disease progression by a proxy variable known as the CAG-Age product (CAP) which is a function of the CAG repeat length and age at study entry. The CAP score, developed using data from 763 individuals in the HD prodrome, exhibits a strong ability to predict diagnosis for the sample described subsequently, based on a receiver operating characteristic analysis (Zhang et al., 2011). Participants can be assigned to Low–Medium–High groups based on CAP using the optimization criteria discussed by Zhang and colleagues. In the current sample, gene-expanded participants were grouped into High (approximately <7.5 years to diagnosis), medium (approximately 7.5–13 years to diagnosis) and Low CAP groups (approximately >13 years to diagnosis) and Controls (non-gene expansive; see Table 1).

**Design**

Data were collected from participants in PREDICT-HD, a 32-site, international, longitudinal, and observational study designed to examine biomarkers (i.e., blood, urine, and imaging) as well as refine clinical markers (cognitive, psychiatric, sensory, and motor) of early disease in people with the gene expansion for HD (Paulsen et al., 2006). Participants provided informed consent for participation. The study was approved by institutional review boards at all study and data-processing sites. Participants are seen annually, with the Towers task administered on odd-numbered visits (years 1, 3, and 5).

**Measures**

**Performance-Based Measures.** The Towers Task, the primary measure of interest in this study, was a computerized measure that included two conditions. For the first condition, Towers 3, participants viewed three vertical pegs. One peg contained a stack of three disks of increasing sizes with the largest disk on the bottom of the stack. The goal was to relocate the stack, in exactly the same configuration, to a different peg while following two rules: (i) only one disk may be moved at a time

<table>
<thead>
<tr>
<th>Table 1. Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP group</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>Med</td>
</tr>
<tr>
<td>High</td>
</tr>
</tbody>
</table>

*Note: SD = standard deviation; groupings pertain to CAP score where High is approximately <7.5 years to diagnosis, Medium is approximately 7.5–13 years to diagnosis, and Low is approximately >13 years to diagnosis and Controls are non-gene expansive family members.*
and (ii) a larger disk may never be stacked on a smaller disk. Participants were instructed to “complete the task in as few moves as possible—do not worry about how fast you finish.” For each attempt at Towers 3, the participant was allowed a maximum of 21 moves (basal) but the task requires at least 7 moves for successful completion (ceiling). If not solved within 21 moves, the screen reset to the original configuration and the next trial began, allowing only one attempt at each trial. Only if participants were able to successfully complete Towers 3 (within the maximum number of moves), participants then completed the second condition, Towers 4, which is a three peg task with four disks and the same rules as described earlier. For each attempt at Towers 4, the participant had a maximum of 45 moves (basal) to solve the task. The participant had to complete at least 15 moves for successful completion (ceiling). Towers 3 and 4 each included four identical trials. The mean number of moves is the average number of moves across four trials. The mean Execution Time equals the average amount of time required between the first to last move, averaged across four trials. Initiation time is the average amount of time required between the presentation of the Towers and the first move, averaged across four trials. Both measures (Execution Time and Initiation Time) are thought to reflect planning. Transformations of timing measures (square roots) were made to normalize the data. Administration of both Towers required approximately 15 min and Towers 3 was administered first.

In order to determine the impact of motor speed on Towers performance, a motor speed measure previously shown to be sensitive to prodromal HD (i.e., finger tapping speed) was included. In this measure, participants are required to tap as quickly as possible for five 10-s intervals using their nondominant hand on a response box interfaced to a computer. Results are reported as the mean and standard deviation of the intertap intervals.

Self- and Informant Report Measures. To determine whether ED as measured by the Towers task correlated with ED in daily life, we administered a modified version of the self- and companion reports of executive problems from the Frontal System Behavior Scale (FrSBe), a normed scale that assesses behavioral disturbances associated with damage to fronto-subcortical circuits (Grace & Malloy, 2001). The modified FrSBe used previously in this study (Duff et al., 2010) contains 24 items that assess the frequency and the level of distress caused by certain behaviors. Frequency response choices range from 1 to 5 (“almost never” to “almost always”), and distress response choices range from 1 to 5 (“not at all distressing” to “extremely distressing or very severe”). For both the participant and companion versions of the modified FrSBe, the total score equals the total sum of the endorsed frequency and distress for each item, with higher scores indicating greater frequency of behaviors indicative of frontal dysfunction and more distress caused by them. In this analysis, we focused on the ED subscale, which evaluates difficulties with working memory, planning, problem-solving, and insight. Towers performance was also examined in relation to mood symptomatology using the Beck Depression Inventory-II, a 21-item self-report rating scale where higher scores indicate greater levels of depression (Beck, Steer, & Brown, 1996).

Brain Volumetric Measures. As part of the PREDICT-HD study, neuroimaging markers of disease progression were also assessed. Magnetic resonance imaging scans were obtained using a standard multi-modal protocol that included an axial 3D volumetric spoiled gradient-echo series (1 × 1 × 1.5 mm voxels) and a dual-echo proton density/\(T_2\) (1 × 1 × 3 mm voxels) series. Thirty sites used General Electric 1.5 T scanners and two sites used Siemens 1.5 T scanners. Scans were processed through an automated procedure implemented in the BRAINS software (Magnotta et al., 2002) re-oriented by stepwise co-registration to a set of template images, centered on the anterior commissure, and resampled to 1 mm resolution. The \(T_1\) image was the final \(T_1\) image, and all images were intensity normalized and homogeneity corrected. A discriminant tissue classification was then performed (Harris et al., 1999) and a brain mask was created using an artificial neural network (ANN) (Powell et al., 2008). Measures of gray and white matter were then completed using the standard Talairach method (Andreasen et al. 1996). ANNs were applied to each scan to measure subcortical structures, including caudate and putamen that were summed to create a measure of total striatal volume. Results of this procedure were visually inspected, and >90% of the scans passed all stages successfully. The most common reason for failure was poor coregistration of the multiple modes. No known variable that was the subject of this report, including gender and HD gene-expansion status, significantly predicted scan failures. Regional volumes were divided by each individual intracranial volume.

Statistical Analysis

The longitudinal analysis was performed using linear mixed effects regression (LMER) (Verbeke & Molenberghs, 2000). LMER analyses were performed using the statistical software program R (v2.13.0) with the function lmer() from the add-on package lme4 (see Douglas, Maechler, & Bolker, 2011). LMER allows the simultaneous assessment of baseline and longitudinal differences by the CAP group (Long, 2012). Repeated measurements on the same subjects induce a correlation structure, which is modeled via random effects (e.g., random intercepts and slopes). Random effect structures were compared using Akaike’s Information Criterion (Akaike, 1973). Models with multiple random effects that were highly correlated (i.e., >.9)
were simplified under the assumption that this reflected instability in the model. The time metric for the analysis was duration, defined as age at time point minus age at study entry. Group intercepts and slopes were allowed to vary by the CAP group, that is, the trajectories were conditional on the CAP group. Maximum likelihood methods were used for estimation, which yields unbiased estimates under the widely applicable assumption that the missing data mechanism is ignorable (Little & Rubin, 2002). The Likelihood Ratio Test (LRT) was used to select the “best” model to report. The candidate models were: (i) No CAP group differences, (ii) CAP group intercept differences, or baseline differences only, and (iii) CAP group intercept and slope differences, allowing for both baseline and longitudinal differences. Covariates were used as control variables in the models. The model adjusted for age at study entry, gender, and education. We added motor speed (finger tapping) to the covariates to determine whether outcomes differed when motor functions were controlled for.

In addition to the longitudinal analysis, a cross-sectional analysis was also performed. The cross-sectional analysis consisted of separate linear regressions (LRs) on cross-sectional baseline data for Towers 3 and 4 outcomes, imaging markers, and behavioral data while controlling for gender, age, and education. From the LR analysis adjusted least-squares means were calculated for each pairing of the Control group and a CAP group. Cross-sectional effect sizes (Cohen’s d) were calculated by dividing the difference in performance between groups by the pooled standard deviation across the two groups being compared (root mean square error). Effect sizes with positive values reflect poorer performance by the prodromal HD groups in comparison with the controls, whereas negative effect-sizes reflect superior performance of the group with prodromal HD compared relative with controls. T-tests were used to statistically verify pair-wise group differences. Partial correlations, controlling for age, education, and gender, were used to examine relationships among CAPs scores (not groups) and imaging and behavioral data. We also examined the baseline characteristics of Towers performance by calculating the percentage of individuals from each group who scored at basal or ceiling levels.

Results

Characteristics of Towers Task Performance

We first examined the characteristics of task performance at baseline. There were no differences between groups on percentage performing at either the basal (7 moves for Towers 3; 15 moves for Towers 4) or ceiling levels (21 moves for Towers 3; 45

![Fig. 1. Towers 4 Execution Time by CAP groups and duration.](image-url)
moves for Towers 4) for number of moves on the first trial of either Towers 3 \( \chi^2(6) = 8.67, p = .1933 \) or Towers 4 \( \chi^2(6) = 9.29, p = .1582 \) (see Figure 2). However, a large percentage of cases and controls completed Towers 3 at the minimum number of moves. The number of participants performing Towers 4 at the minimum number of moves was smaller.

Table 2. Summary of intercepts and slopes for all measures (controlling for age, gender, and education)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Intercepts</th>
<th>Slopes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Low</td>
</tr>
<tr>
<td>Towers 3 Average Moves</td>
<td>9.36**</td>
<td>8.88</td>
</tr>
<tr>
<td>Towers 3 Average Initiation</td>
<td>2.37***</td>
<td>2.45</td>
</tr>
<tr>
<td>Towers 3 Average Execution</td>
<td>4.62***</td>
<td>4.57</td>
</tr>
<tr>
<td>Towers 3 Average Total Time</td>
<td>5.27***</td>
<td>5.27</td>
</tr>
<tr>
<td>Towers 4 Average Moves</td>
<td>26.07***</td>
<td>26.55</td>
</tr>
<tr>
<td>Towers 4 Average Initiation</td>
<td>2.09***</td>
<td>2.29*</td>
</tr>
<tr>
<td>Towers 4 Average Execution</td>
<td>7.54***</td>
<td>8.09*</td>
</tr>
<tr>
<td>Towers 4 Average Total Time</td>
<td>7.86***</td>
<td>8.46*</td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01, ***p < .001.

Fig. 2. Percentage of control and high CAP groups performing at either the basal or ceiling level for number of moves per trial on towers 3 and towers 4.
Covarying for age, education, and gender, statistically significant baseline (i.e., intercept) differences among groups (i.e., Low, Medium, and High CAP groups and Controls) were found for all outcomes: Mean number of moves for both Towers 3 \([\text{LRT} (\chi^2(3)) = 22.50, p < .0001]\) and Towers 4 \([\text{LRT} (\chi^2(3)) = 26.26, p < .0001]\), Initiation Time for Towers 3 \([\text{LRT} (\chi^2(3)) = 16.36, p = .001]\) and Towers 4 \([\text{LRT} (\chi^2(3)) = 31.36, p < .0001]\), and for Execution Time on Towers 3 \([\text{LRT} (\chi^2(3)) = 72.65, p < .0001]\) and Towers 4 \([\text{LRT} (\chi^2(3)) = 101.52, p < .0001]\). In terms of the cross-sectional effect sizes, differences were most evident when comparing the High group with controls, High with Low, and High with Medium (see Tables 2 and 3). Effect sizes amongst groups were larger for Execution Time, particularly on Towers 4 versus Towers 3. The only significant group difference between the Low and Control groups on Towers performance was evident on Towers 4 Execution Time (Cohen’s \(d = 0.23\)).

No longitudinal differences (i.e., slope) were found amongst groups for Mean Moves or Initiation Time but longitudinal differences were seen on Towers 4 \([\text{LRT}(\chi^2(3)) = 8.11, p = .0438]\). In fact, all groups appear to improve completion of the task over time. Figure 1 displays group differences for the most sensitive measure (Execution Time on Towers 4; sensitivity described further subsequently) at baseline and over follow-up by group. The Low CAP group improves most (slope = −.24), followed by the Control and Medium groups (slopes = −.15 and −.16 respectively), and finally the High group, with the least improvement over time (slope = −.04).

Table 4 provides normative data for gene-expanded individuals performing below the control mean by 1, 1.5, and 2 standard deviations. Approximately 36% of individuals in the High group perform one or more standard deviations below the control mean. Approximately 23% of individuals in the Medium group and 13% of individuals in the Low group perform one or more standard deviations below the Control mean.

### Analyses of Towers Task, Brain Metrics, and Daily Functioning

Baseline performance by gene-expanded individuals on the Towers task was related to both frontal white matter and striatal volume. The mean striatal volume was 0.00962 \((SD = 0.00175)\) and mean frontal white matter was 0.12962 \((SD = 0.01370)\). Striatal volume was significantly negatively correlated with Execution Time for both Towers: \(r = -.16, p = .0007\) for Towers 3, and \(r = -.20, p < .0001\) for Towers 4. Striatal volume was also significantly related to the mean number of moves: \(r = -.17, p = .0003\) for Towers 3, and \(r = -.14, p = .0026\) for Towers 4. Frontal white matter was significantly correlated with both Towers: \(r = -.12, p = .007\) for Towers 3, and \(r = -.19, p < .0001\) for Towers 4.

We also examined the relationship between Towers performance to executive functioning in daily life and mood symptomatology. No relationship was observed between self-report of ED on the Frontal Systems Behavioral Scale (FrSBe) in gene-expanded individuals and performance on Mean Number of Moves, Initiation Time, or Execution Time for either Towers task (see Table 5). However, there was a significant correlation between companion endorsement of ED on the FrSBe and Execution Time for Towers 4, \(r = .15, p = .0008\).

#### Table 3. Baseline effect sizes (Cohen’s \(d\)) between CAP groups (controlling for age, gender and education)

<table>
<thead>
<tr>
<th></th>
<th>Mean Number of Moves</th>
<th>Initiation Time</th>
<th>Execution Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tower 3</td>
<td>Tower 4</td>
<td>Tower 3</td>
</tr>
<tr>
<td>High versus Control</td>
<td>0.21*</td>
<td>0.33***</td>
<td>0.31***</td>
</tr>
<tr>
<td>Med versus Control</td>
<td>0.05</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>Low versus Control</td>
<td>−0.16</td>
<td>0.10</td>
<td>−0.02</td>
</tr>
<tr>
<td>High versus Low</td>
<td>0.42***</td>
<td>0.25*</td>
<td>0.29**</td>
</tr>
<tr>
<td>High versus Med</td>
<td>0.17*</td>
<td>0.24**</td>
<td>0.17*</td>
</tr>
</tbody>
</table>

Note: *\(p < .05\), **\(p < .01\), ***\(p < .001\).

#### Table 4. Percentage of cases −1, −1.5 and −2 \(SD\) below the Control mean for execution time on Towers 4

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>44.16%</td>
<td>12.69%</td>
<td>6.09%</td>
<td>5.08%</td>
<td>197</td>
</tr>
<tr>
<td>Medium</td>
<td>54.68%</td>
<td>23.02%</td>
<td>13.67%</td>
<td>8.63%</td>
<td>278</td>
</tr>
<tr>
<td>High</td>
<td>66.99%</td>
<td>35.95%</td>
<td>26.14%</td>
<td>18.63%</td>
<td>306</td>
</tr>
</tbody>
</table>
Table 5. Partial correlation matrix for gene-expanded individuals (n = 469)

<table>
<thead>
<tr>
<th></th>
<th>T3 Moves</th>
<th>T4 Moves</th>
<th>T3 Initiation Time</th>
<th>T4 Initiation Time</th>
<th>T3 Execution Time</th>
<th>T4 Execution Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP Score</td>
<td>0.16**</td>
<td>0.14**</td>
<td>0.05</td>
<td>0.02</td>
<td>0.19***</td>
<td>0.23***</td>
</tr>
<tr>
<td>BDI-II</td>
<td>0.07</td>
<td>0.03</td>
<td>−0.05</td>
<td>0.01</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>Striatal Volume</td>
<td>−0.17**</td>
<td>−0.14**</td>
<td>−0.02</td>
<td>0.00</td>
<td>−0.16**</td>
<td>−0.20***</td>
</tr>
<tr>
<td>Frontal White Matter Volume</td>
<td>−0.05</td>
<td>−0.08</td>
<td>−0.00</td>
<td>−0.10</td>
<td>−0.12**</td>
<td>−0.19***</td>
</tr>
<tr>
<td>Self Report FrSBe Executive</td>
<td>0.04</td>
<td>0.05</td>
<td>−0.10</td>
<td>−0.02</td>
<td>−0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Companion FrSBe Executive</td>
<td>0.07</td>
<td>0.10</td>
<td>−0.01</td>
<td>0.07</td>
<td>0.07</td>
<td>0.15**</td>
</tr>
<tr>
<td>T3 Moves</td>
<td>.31***</td>
<td></td>
<td>−0.00</td>
<td>−0.00</td>
<td>0.71***</td>
<td>0.23***</td>
</tr>
<tr>
<td>T4 Moves</td>
<td>−0.09</td>
<td></td>
<td>0.05</td>
<td>0.24***</td>
<td>0.57***</td>
<td></td>
</tr>
<tr>
<td>T3 Initiation Time</td>
<td></td>
<td></td>
<td></td>
<td>0.20***</td>
<td>0.16**</td>
<td></td>
</tr>
<tr>
<td>T4 Initiation Time</td>
<td></td>
<td></td>
<td></td>
<td>0.19***</td>
<td>0.37***</td>
<td></td>
</tr>
<tr>
<td>T3 Execution Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50***</td>
</tr>
</tbody>
</table>

Note: T3 = Towers 3, T4 = Towers 4, BDI-II = Beck Depression Inventory II, FrSBe = Frontal Systems Behavioral Scale, **p < .01, ***p < .001.

Performance on the Towers task (Mean Number of Moves, Initiation Time, and Execution Time) was unrelated to symptoms of depression, as measured by the Beck Depression Inventory-II in gene-expanded individuals.

Discussion

Performance on the Towers Task, a measure of executive function, was significantly related to cumulative genetic toxicity (i.e., disease progression at time of study entry) in participants in the HD prodrome. Those closer to estimated motor diagnosis at the time of study entry required more time and more moves to complete the Towers Task, indicating that they had greater difficulty efficiently holding and making use of a plan in working memory compared with individuals farther from estimated diagnosis and controls. Interestingly, the time required to complete the task was more sensitive to baseline progression (CAP group) compared with other outcomes including Mean Number of Moves and Initiation Time. The negative relationship between Execution Time (but not Mean Number of Moves) and CAP group remained after controlling for motor speed, indicating that the Towers Task is providing information about executive functioning beyond simply reduced motor speed. Approximately 36% of individuals in the High group (i.e., those approximately <7.5 years to diagnosis), 23% of participants in the Medium group (approximately 7.5–13 years to diagnosis), and 13% of individuals in the Low group (approximately >13 years to diagnosis) are performing one or more standard deviations below the control mean on Towers 4 Execution Time. In addition, approximately 19% of individuals in the High group performed two or more standard deviations below controls on Towers 4 Execution Time, suggesting that a subset of those with greater disease progression is exhibiting significant ED before manifest HD. Previous studies have shown large effect sizes between symptomatic HD and controls on Towers tasks (Hanes et al. 1996). Compared with previous studies, cross-sectional effect sizes were in the medium range amongst groups for Towers 4 Execution Time in comparison with small effect sizes for other Towers outcomes. Towers 4 Execution Time was the only outcome which showed group differences on the longitudinal follow-up. Towers 4 Execution Time was also modestly correlated with companion report of ED in daily life while other Towers outcomes were not, suggesting that Towers 4 Execution Time may have clinical utility. Furthermore, Execution Time was negatively correlated with both frontal white matter volume and striatal volume while increased Mean Number of Moves was related to only smaller striatal volume which indicates that the different Towers outcomes may be differentially enlisting brain regions to perform the task. Accurate performance (e.g., Mean Moves) is related to striatal volume whereas efficient working memory (e.g., Execution Time) is related to both striatal and frontal white matter volume.

Most outcomes on the Towers Task were not associated with increasing disease pathology at longitudinal follow-up, which may be the result of task limitations: (i) ceiling effects on Towers 3 given that a large percentage of participants completed Towers 3 in the minimum number of moves and (ii) practice effects in that all groups improved over follow-up. It is important to note, however, that for Towers 4 Execution Time (see Figure 1), the High CAP group exhibits the least improvement over time. The lack of improvement on the task may be a more useful marker in those closer to disease onset rather than frank decline and previous research in manifest HD indicates that a lack of practice effects is associated with poorer cognitive outcomes (Duff et al., 2007). Nonetheless, the second exposure to a Towers Task may be fundamentally different from the first where a novel strategy must be created and implemented. Although it has been reported that the first exposure “spoils” the validity of subsequent administration’s measure of executive functioning, our data suggest that it may be that the spoiling effect is attenuated by increasing the working memory component of the task, as few individuals (either controls or cases) performed at the basal level on Towers 4 despite being previously being exposed to Towers 3. Towers 4 (particularly...
Execution Time) was also more sensitive in detecting the group differences and because of this has clinical utility (normative data provided in Table 4).

Given that motor performance can be a confound in HD gene-expanded carriers on some cognitive measures, it would appear that moves required would be less sensitive to motor impairments than the time required to solve the task. Our contrary finding, however, is consistent with prior work in early HD patients indicating impaired planning despite intact decision-making (Watkins et al., 2000), given that Execution Time is hypothesized to be a measure of efficient preplanning (Owen, Downes, Sahakian, Polkey, & Robbins, 1990).

Gene-expanded carriers were accurate but less efficient at planning and executing the required sequence of moves. Performance also correlated negatively with striatal and frontal white matter volume, consistent with the striatum’s key role in efficient planning and problem solving in this task (Sullivan et al., 2009) and recent research associating ED in early HD with striatal atrophy (Peinemann et al., 2005). Performance on Towers 3 and Towers 4 were correlated ($r = .31$ for Mean Moves, $r = .50$ for Execution Time), which indicates that increasing the working memory load of the task had a significant effect on performance; the percentage of individuals performing the task in the fewest trials possible (basal level) is much higher for Towers 3 than for Towers 4 (see Figure 2). Additionally, Towers 4 was more sensitive at detecting cross-sectional differences between groups compared with Towers 3 (see effect sizes in Table 3), indicating that the increased working memory component was more challenging to those with a higher CAP score. More specifically, there was a statistically significant effect size when comparing Execution Time on Towers 4 in the Low versus Control group, but no significant difference between Low and Control on Execution Time for Towers 3.

Although some gene-expanded individuals present symptoms of ED prior to and following motor diagnosis (Duff et al., 2010), heterogeneity in prodromal cognitive symptoms has been found (Rosas et al., 2008). We found a small correlation between companion report of ED and Towers performance. This indicates that the task may be more useful in a subset of people who are already showing symptoms of ED in daily life. It also highlights the potential dissociations between subjective self-report and objective performances across the spectrum of HD (Duff et al., 2010).

Although the Towers task was unable to distinguish among the CAP group in terms of change over time, these results have provided hypotheses about early ED and directions for future work. ED is measurable in HD gene expansion carriers’ years before motor diagnosis. Lack of improvement on follow-up rather than frank decline may be used as a proxy for disease progression, especially in the prodromal period. Relating performance in gene expanded individuals to age-matched non-affected family members is also informative. Including more cognitively challenging Towers Tasks (five or six disks) in those farther from estimated diagnosis may detect more subtle deficits and may potentially overcome task familiarity to be more sensitive to measuring change over time.

Although this study examined a large group of prodromal HD gene expansion carriers, it should be noted that the percentage of at-risk individuals who decide to obtain genetic testing is relatively low so these findings may only apply to this resourceful and self-selected group (Decruyenaere et al., 1995). Another limitation of this study is that although we examined Towers performance in relation to self- and companion report of ED in daily life, we did not look at functional changes (e.g., declines in activities of daily living) which can affect employment and independent living. Furthermore, executive functioning is a broad and complex cognitive construct and this study is limited by looking at only one aspect of problem-solving skills.

In conclusion, these findings indicate that ED in prodromal HD varies by baseline progression status. The total time required to complete the Towers Task is a useful indicator of executive functioning that adds information beyond differences in motor control and will likely have utilitarian value in clinical settings (norms provided in Table 4).

**Funding**

This work was supported by the National Institutes of Health [2RO1 NS0040068]; National Institute of Neurological Disorders and Stroke [NS40068]; and CHDI Foundation, Inc.

**Conflict of Interest**

None declared.

**Acknowledgements**

We thank the PREDICT-HD sites, the study participants, the National Research Roster for Huntington Disease Patients and Families, and the Huntington Disease Society of America. See the Appendix for the full list of PREDICT-HD Investigators, Coordinators, Motor Raters, Cognitive Raters. We thank the following experts who participated in teleconferences and
provided feedback on earlier drafts of this project: Kevin Duff, Leigh Beglinger, David J. Moser, Kelly C. Rowe, Megan Smith, Julie Stout, Deborah Harrington, Gabriel Castillo, Jessica Morison, and Jason Reed, Michael Diaz, Ian Dobbins, Tamara Hershey, Erin Foster, Deborah Moore, Holly Westervelt, Jennifer Davis, Geoff Tremont, Gloria Wenman, Danielle Theriault, Carissa Gehl, Kirsty Matheson, Karen Siedlecki, Marleen Van Walsem, Susan Bonner, Greg Elias, Mary Gever, Rachel Bernier, Melanie Faust, Noelle Carlozzi, Nellie Georgiou-Karistianis, and Herwig Lange. This task was developed in the Indiana Cognitive Research lab of Julie Stout and we appreciate her early contribution to the development of this task.

Appendix: PREDICT-HD Investigators, Coordinators, Motor Raters, Cognitive Raters

Active: September 2010–August 2011

Thomas Wassink, MD, Stephen Cross, BA, Mycah Kimble, BA, Patricia Ryan, MSW, LISW, MA, Jessica Wood, MD, PhD, Eric A. Epping, MD, PhD, and Leigh J. Beglinger, PhD (University of Iowa, Iowa City, IA, USA); Edmond Chiu, MD, Olga Yastrubetskaya, PhD, Joy Preston, Anita Goh, D.Psych, Chatrushka Fonseka, Stephanie Antonopoulos and Samantha Loi (St. Vincent’s Hospital, The University of Melbourne, Kew, Vic., Australia); Phyllis Chua, MD, and Angela Komiti, BS, MA (The University of Melbourne, Royal Melbourne Hospital, Melbourne, Australia); Lynn Raymond, MD, PhD, Rachelle Dar Santos, BSc, Kimberley Carter, BSc, and Joji Decolongon, MSC, CCRP (University of British Columbia, Vancouver, BC, Canada); Adam Rosenblatt, MD, Christopher A. Ross, MD, PhD, Barnett Shpritz, BS, MA, OD, Nadine Yoriotomo, RN and Claire Welsh (Johns Hopkins University, Baltimore, MD, USA); William M. Mallonee, MD, Greg Suter, BA, and Judy Addison (Hereditary Neurological Disease Centre, Wichita, KS, USA); Ali Samii, MD, and Alma Macaraeg, BS (University of Washington and VA Puget Sound Health Care System, Seattle, WA, USA); Randi Jones, PhD, Cathy Wood-Siverio, MS, Stewart A. Factor, DO, and Claudia Testa, MD, PhD (Emory University School of Medicine, Atlanta, GA, USA); Roger A. Barker, BA, MBBS, MRCP, Sarah Mason, BSc, Anna Goodman, PhD, and Anna DiPietro (Cambridge Centre for Brain Repair, Cambridge, UK); Elizabeth McCusker, MD, Jane Griffith, RN, Clement Loy, MD, and David Gunn, BS (Westmead Hospital, Sydney, Australia); Bernhard G. Landwehrmeyer, MD, Michael Orth MD, PhD, Sigurd Sußmuth, RN, Katrin Barth, RN, and Sonja Trautmann, RN (University of Ulm, Ulm, Germany);

Kimberly Quaid, PhD, Melissa Wesson, MS, and Joanne Wojcieszek, MD (Indiana University School of Medicine, Indianapolis, IN);

Mark Guttmann, MD, Alanna Sheinberg, BA, and Irita Karmalkar, BSc (Centre for Addiction and Mental Health, University of Toronto, Markham, Ont., Canada);

Susan Perlman, MD, Brian Clemente, and Arik Johnson, PsyD (University of California, Los Angeles Medical Center, Los Angeles, CA, USA);

Michael D. Geschwind, MD, PhD, Jon Gooblar, BA, and Gail Kang, MD (University of California San Francisco, CA, USA);

Tom Warner, MD, PhD, Maggie Burrows, RN, BA, Marianne Novak, MD, Thomasin Andrews, MD, BSC, MRCP, Elisabeth Rosser, MBBS, FRCP, and Sarah Tabrizi, MD, PhD (National Hospital for Neurology and Neurosurgery, London, UK);

Anne Rosser, MD, PhD, MRCP, Kathy Price, RN, and Sarah Hunt, BSc (Cardiff University, Cardiff, Wales, UK);

Frederick Marshall, MD, Amy Chesire, LCSW-R, MSG, Mary Wodarski, BA, and Charlyne Hickey, RN, MS (University of Rochester, Rochester, New York, USA);

Oksana Suchowersky, MD, FRCP, Sarah Furtado, MD, PhD, FRCP, and Mary Lou Klimek, RN, BN, MA (University of Calgary, Calgary, AB, Canada);

Peter Panegyres, MB, BS, PhD, Joseph Lee, and Steve Andrew (Neurosciences Unit, Graylands, Selby-Lemnos & Special Care Health Services, Perth, Australia);

Joel Perlmutter, MD, Stacey Barton, MSW, LCSW, and Amy Schmidt (Washington University, St. Louis, MO, USA); Zosia Miedzybrodzka, MD, PhD, Daniela Rae, RN, and Mariella D’Alessandro, PhD (Clinical Genetics Centre, Aberdeen, Scotland, UK);

David Craufurd, MD, Ruth Fullam, BSC, Judith Bek, PhD, and Elizabeth Howard, MD (University of Manchester, Manchester, UK);

Pietro Mazzoni, MD, PhD, Karen Marder, MD, MPH, and Paula Wasserman, MA (Columbia University Medical Center, New York, NY, USA);

Rajeev Kumar, MD and Diane Erickson, RN (Colorado Neurological Institute, Englewood, CO, USA);

Vicki Wheelock, MD, Terry Tempkin, RNC, MSN, Lisa Kjer, MSW, and Kathleen Baynes, PhD (University of California Davis, Sacramento, CA, USA);
Joseph Jankovic, MD, Christine Hunter, RN, CCRC, and William Ondo, MD (Baylor College of Medicine, Houston, TX, USA);
Wayne Martin, MD, Pamela King, BScN, RN, Marguerite Wieler and Satwinder Sran, BSC (University of Alberta, Edmonton, AB, Canada);
Anwar Ahmed, PhD, Stephen Rao, PhD, Christine Reece, BS, Janice Zimbelman, PhD, PT, Alexandra Bea, BA, Emily Newman, BA, Alex Bura, BA, Lyla Mourany, and Juliet Schulz (Cleveland Clinic Foundation, Cleveland, OH, USA).

Executive Committee

Jane Paulsen, PhD, Principal Investigator, Eric A. Epping, MD, PhD, Megan Smith, PhD, Jeffrey D. Long, PhD, Hans Johnson, Ph.D., Jeremy Bockholdt, Kelsey Montross.

Core Sections

Brain: Jean Paul Vonsattell, PhD (Chair), and Carol Moskowitz, ANP, MS (Columbia University Medical Center).
Cognitive: Deborah Harrington, PhD (Chair), Gabriel Castillo, BS, Jessica Morison, BS, and Jason Reed, BS (University of California, San Diego), Michael Diaz, PhD, Ian Dobbins, PhD, Tamara Hershey, PhD, Erin Foster, OTD, and Deborah Moore, BA (Washington University Cognitive Science Battery Development); Holly Westervelt, PhD (Chair, Quality Control and Training, Alpert Medical School of Brown University), Jennifer Davis, PhD, and Geoff Tremont, PhD, MS (Scientific Consultants, Alpert Medical School of Brown University); Megan Smith, PhD (Chair, Administration), David J. Moser, PhD.
Functional: Janet Williams, PhD (Chair), Nancy Downing, RN, PhD, Joan Laing, PhD, Kristine Rees, BA, Michelle Harrel, BS, and Stacie Vik, BA (University of Iowa); Rebecca Ready, PhD (University of Massachusetts); Anthony Vaccarino, PhD (Ontario Cancer Biomarker Network); Sarah Farias, PhD (University of California, Davis); Noelle Carlozzi, PhD (University of Michigan); and Carissa Gehl, PhD (VA Medical Center, Iowa City, IA).
Imaging: Hans Johnson, PhD, Elizabeth Aylward, PhD (Chair, Seattle Children’s Research Institute). Eric Axelson, BSE (University of Iowa). Christopher A. Ross (Chair), MD, PhD, Vincent A. Magnotta, PhD (Chair, University of Iowa).
Motor: Kevin Biglan, MD (Chair) (University of Rochester); Karen Marder, MD (Columbia University); Jody Corey-Bloom, MD, PhD (University of California, San Diego); Michael Geschwind, MD, PhD (University of California, San Francisco); Ralf Reilmann, MD and Zerka Unds (Muenster, Germany).
Psychiatric: Eric A. Epping, MD, PhD (Chair), Nancy Downing, RN, PhD, Jess Fiedorowicz, MD, Robert Robinson, MD, Megan Smith, MD.
Statistics: Jeffrey D. Long, PhD, Ji-In Kim, PhD, James A. Mills, MS, Ying Zhang, PhD, Dawei Liu, PhD, Wenjing Lu, and Spencer Lourens (University of Iowa).
Ethics: Cheryl Erwin, JD, PhD, (Chair, McGovern Center for Health, Humanities and the Human Spirit); Eric A. Epping, MD, PhD Janet Williams, PhD, James Mills, MS, Martha Nance, MD (University of Minnesota); and Lisa Hughes, MEd (University of Texas Medical School at Houston).
IT/management: Hans Johnson, PhD (Chair), R.J. Connell, BS, Karen Pease, BS, Ben Rogers, BA, BSCS, Jim Smith, AS, Shuhua Wu, MCS, Roland Zschiegner, Erin Carney, Bill McKirgan, Mark Scully, and Ryan Wyse (University of Iowa); Jeremy Bockholt (AMBIGroup).

Program Management

Administrative: Chris Werling-Witkoske (Chair), Stacie Vik, BA, Karla Anderson, BS, Sean Thompson, BA, Brittany Lichty, BA, Leann Davis, Craig Stout (University of Iowa).
Financial: Steve Blanchard, MSHA, Kelsey Montross, BA, and Phil Danzer (University of Iowa).

References
