MONASH University

# Remote Assessment of Cognition, Sleep, and Physical Activity in Early Huntington's Disease

Brendan John McLaren
BBioSc (Hons), GradDipPsych, BSc (Hons)

A thesis submitted for the degree of

***Doctor of Psychology (Clinical Neuropsychology)***

Monash University in 2020

School of Psychological Sciences

Faculty of Medicine, Nursing, and Health Sciences

Monash University

Australia

## Copyright notice

## Table of Contents

## List of Prepared Manuscripts and Presentations

This thesis includes the following manuscript accepted for publication during candidature:

**McLaren, B**., Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E.-C., Murray, N. W. G., Loy, C., Bellgrove, M. A., & Stout, J. C. (2020). Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *Journal of Neurology*.

This thesis includes the following manuscripts submitted for peer-review during candidature:

**McLaren, B.,** Romeu, R. J., Andrews, S. C., Bellgrove, M. A., & Stout, J. C. (2020). Diffusion Modeling of a Two-choice Decision-making Task in Premanifest and Manifest Huntington's Disease.

**McLaren, B.,** Drummond, S. P. A., Glikmann-Johnston, Y., Loy, C., Bellgrove, M. A., Stout, J. C., & Andrews, S. C. (2020). Greater time in bed and less physical activity associate with poorer cognitive functioning performance in Huntington's disease.

Parts of this thesis were presented at the following local symposia and national conferences:

**McLaren, B.,** Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E., Bellgrove, M. A., Loy, C., Drummond, S. P. A., & Stout, J. C. Feasibility of Mobile Clinical and Sensor-Based Outcomes in Huntington's Disease. National Institute for Dementia Research Australian Dementia Forum 2017,
Melbourne, Australia.

**McLaren, B.,** Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E., Bellgrove, M. A., Loy, C., Drummond, S. P. A., & Stout, J. C. Validation of a Novel Smartphone-Based Cognitive

Assessment Tool Designed for Remote Assessment of Cognitive Functioning in Huntington's Disease. 17th National Conference of Emerging Researchers in Ageing 2018, Melbourne, Australia.

Parts of this thesis were presented at the following international conferences:

**McLaren, B.,** Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E., Bellgrove, M. A., Loy, C., Drummond, S. P. A., & Stout, J. C. Feasibility of Mobile Clinical and Sensor-Based Outcomes in Huntington's Disease. 12th Annual Huntington's Disease Therapeutics Conference (CHDI Foundation, inc) 2017, St Julian's, Malta

**McLaren, B**., Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E., Bellgrove, M. A., Loy, C., Drummond, S. P. A., & Stout, J. C. Validation of a Novel Smartphone-Based Cognitive Assessment Tool Designed for Remote Assessment of Cognitive Functioning in Huntington's Disease. European Huntington's Disease Network Plenary Meeting 2018**,** Vienna, Austria

## List of Tables

**Chapter Five**

## List of Figures

### Chapter Two



### Chapter Three



### Chapter Four

## Abbreviations and Symbols

| | |
|---|---|
| ANOVA | Analysis of Variance |
| CAG | cytosine-adenine-guanine |
| *d* | Cohen's *d* estimate of effect size |
| DMSAQ | Daily Mood Sleep and Activity Questionnaire |
| ECG | Electrocardiogram |
| EEG | Electroencephalography |
| *F* | F-test |
| HADS | Hospital Anxiety and Depression Scale |
| HC | Healthy control |
| HD | Huntington's disease |
| IPAQ | International Physical Activity Questionnaire |
| KSS | Karoninska Sleepiness Scale |
| ms | milliseconds |
| *n* | number |
| *p* | *p*-value (significance test) |
| PSG | Polysomnography |
| REM | Rapid Eye Movement |
| RDM | Ratcliff Diffusion Model |
| SCOPA-SLEEP | SCale of Outcomes in PArkinsons disease |
| SDMT | Symbol Digit Modalities Test |
| TFC | Total Functioning Capacity |
| TMS | Total Motor Score |
| UHDRS | Unified Huntington's Disease Rating Scale |
| $\eta_p^2$ | Partial eta-squared estimate of effect size |

**Abstract**

Huntington's disease (HD) is an inherited, neurodegenerative disease with typical onset in middle adulthood. HD causes progressive cognitive, motor, and psychiatric symptoms. Cognitive decline is an essential feature of HD, which emerges early in the disease and leads to devastating effects on functioning and quality of life. Due to the importance of cognition, research has characterised the course of cognitive symptom progression in HD. Until now, longitudinal cognition research in HD has been limited to clinician-based assessment at infrequent (e.g., annual) timepoints. Clinic-based assessments, however, preclude understanding of how people are functioning in their natural (i.e., home) environment, or in relation to daily sleep and exercise levels. In this thesis research, I examined remote, smartphone-based cognitive assessment as a solution for enabling rapid, inexpensive, frequent, and convenient collection of cognitive data in HD. Specifically, we developed a novel 'bring your own' smartphone-based cognitive assessment with ongoing monitoring of sleep and physical activity, provided by an inexpensive wearable device, to capture dynamic interactions between cognition, sleep, and physical activity. Our system was an exclusively remote means of collecting data on the day to day cognitive functioning, sleep, and physical activity habits of people with premanifest- and early manifest-HD. The research showed that the mobile app-activity monitoring system we developed was feasible and acceptable by people with the expanded HD repeat. The data it yielded were reliable and showed known groups validity, distinguishing between HD and neurologically healthy controls. The cognitive measures were also sensitive to severity, separating performance of manifest-HD and premanifest-HD participants and controls. We found associations between the sleep and physical activity habits of participants and their cognitive functioning. We also applied diffusion modelling to cognitive decision data collected in the app and found that we could augment distinctions

between the manifest- and premanifest-HD groups. Overall, we showed for the first time in HD, that a mobile app could be designed and implemented without ever meeting participants in person or supplying formal training, and that the app could generate reliable, valid data on cognition, sleep, and physical activity. Remote cognitive and lifestyle assessment is an important enabler for the collection of large datasets needed for discovering genetic modifiers of disease onset and conducting post-marketing surveillance of treatments.

## General Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes one original paper published in a peer reviewed journal and two submitted publications. The core theme of the thesis is remote assessment of cognition, sleep, and physical activity in Huntington's disease. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Psychological Sciences under the supervision of Professor Julie Stout.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.
In the case of Chapter Three, Four, and Five my contribution to the work involved the following:

| **Thesis Chapter** | **Publication Title** | **Status** | **Nature and % of student contribution** | **Co-author name(s) Nature and % of co-author's contribution** | **Co-author(s), Monash student Y/N** |
|---|---|---|---|---|---|
| Three | Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease | *Accepted* | 60% contribution by candidate. This included formulation of experimental design, data collection, data preparation and analysis, and writing manuscript | 1) Sophie C. Andrews, project conception, statistical analysis review, manuscript review, 10%<br>2) Yifat Glikmann-Johnston, statistical analysis review, | No<br><br>No |

| | | | | | |
|---|---|---|---|---|---|
| | | | | manuscript review, 5%<br>3) Emily-Clare Mercieca, recruitment of participants, manuscript review, 2.5%<br>4) Nicholas W. G. Murray, recruitment of participants, manuscript review, 2.5%<br>5) Clement Loy, statistical analysis review, manuscript review, 5%<br>6) Mark A. Bellgrove, statistical analysis review, manuscript review, 5%<br>7) Julie C. Stout, project conception, Statistical analysis review, manuscript review, 10% | No<br>No<br>No<br>No<br>No |
| Four | Diffusion Modeling of a Two-choice Decision-making Task in Premanifest and Manifest Huntington's Disease | Submitted | 65% contribution by candidate. This included formulation of experimental design, data collection, data preparation and analysis, and writing manuscript | 1) Ricardo J. Romeu, data analysis (Bayesian modeling), manuscript preparation, and review, 10%<br>2) Sophie C. Andrews, project conception, statistical analysis review, manuscript review, 10%<br>3) Mark A. Bellgrove, statistical analysis | No<br>No<br>No |

| | | | | | |
|---|---|---|---|---|---|
| | | | | review, manuscript review, 5%<br>4) Julie C. Stout, project conception, Statistical analysis review, manuscript review, 10% | No |
| Five | Greater time in bed and less physical activity associate with poorer cognitive functioning performance in Huntington's disease | Submitted | 65% contribution by candidate. This included formulation of experimental design, data collection, data preparation and analysis, and writing manuscript | 1) Sean P. A. Drummond, project conception, statistical analysis review, manuscript review, 5%<br>2) Yifat Glikmann-Johnston, statistical analysis review, manuscript review, 5%<br>3) Clement Loy, statistical analysis review, manuscript review, 2.5%<br>4) Mark A. Bellgrove, statistical analysis review, manuscript review, 2.5%<br>5) Julie C. Stout, project conception, Statistical analysis review, manuscript review, 10%<br>6) Sophie C. Andrews, project conception, statistical analysis review, manuscript review, 10% | No<br>No<br>No<br>No<br>No<br>No |

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

**Student name:** Brendan John McLaren

**Student signature:**  **Date:** 11/09/2020

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

**Main Supervisor name:** Professor Julie Stout

**Main Supervisor signature:**  **Date:** 11/09/2020

**Acknowledgements**

Thank you. Thank you to everyone that helped me get to this point. It has been a long and trying, and rewarding and worthwhile of course(!), process to get to this point. There are so many people that helped in small ways and large. I don't have the most amazing memory, just ask my fiancé, but one thing I don't forget is the people that help me in my life. I will always remember each person who made it possible for me to be sitting at my computer, at 06:35am on the morning that this thesis is due, writing this acknowledgement. I appreciate every one of you. Of course, it is expected that one specifically acknowledges people by name in this section and that will come next. I confess though, that I find it impossible to order my list of specific acknowledgements in a way that makes complete sense. How do you rank these things? So, the following is in no specific order.

To my fiancé' Claire. My life is better for meeting you. I love you. You make me happy, make me laugh, make me a better person, and know how to remove my cares and worries. I look forward to our future together and the options and stability in life that I hope my degree will afford us.

To my dearly beloved, and sadly lost from this world, mum, Robyn. I miss you and wish you were here in person instead of just in my heart. This is for you. To my amazing and ever supportive dad, Robin. There could be no better parents than you and mum. I would not be writing the acknowledgment to a Doctoral thesis if it wasn't for you. I will never truly be able to repay you for all you have done but I will try. To my sister Meagan. Thank you for always believing in and being proud of your little brother.

To my supervisors, Julie, Sophie, and Mark. We did it! It sure took a long while, but we did it! Julie, counting my honours degree and the doctorate, we have worked together for over six years. In that time, I had two major surgeries (you got to have surgery too and unfortunately, your scar is way cooler than mine), lost confidence numerous times, lost loved

ones, and gained a bunch of stuff too. You have been the perfect support and guide through all of that. I am always in awe of just how passionate and indefatigable you are. Sophie, you truly are a delight. Thank you for your unending support and patience. It didn't matter that you moved cities twice during my Doctorate. You were never not there, never not encouraging, and never lost patience. Mark, I am grateful for your prompt responses to whatever I needed and to your frank and no-nonsense nature. A nervous person like me really needed that.

Ricky, thank you for agreeing to join on a project with some guy from the other side of the world that you had never heard of. Chapter Four doesn't happen without your modeling skills and knowledge. I'm glad that we made that paper happen and glad that I made a friend during the process. Here's to meeting in Denver one day!

I have so many friends I can thank, but I want to add a special mention for Jung Kim. Mate, thank you for your support and encouragement. You have a heart. A big one that cares a lot. I can't express all the ways you have helped. But you have, and it has meant a lot.

To my faithful support cat, Cherubael. Thank you for hugs, sandpaper kisses, ambushes, wrestles, and most importantly for sitting by my side every day for six plus months while I completed this thesis in COVID enforced lockdown. You are a naughty boy, but I love you. And, to Cherubael's brother, Schmoo. Thank you for being the most handsome and sweet creature and for keeping Claire company while Cherubael and I slaved away on the thesis.

My final thank you is to all of my participants and to the members of the HD community who assisted with the development of my research protocol. You guys are the real heroes. I hope the research contained in these pages will benefit you.

Thank you again, everyone. Let's see what's next, huh?

# Chapter 1: General Introduction and Thesis Overview

## 1.1 General Overview of the Themes Covered in the Introduction and Description of the Thesis Structure

Huntington's disease (HD) is an autosomal dominant neurodegenerative disease, which has a typical onset in middle adulthood and results in progressive cognitive, motor, and psychiatric symptoms, and eventually death (F. O. Walker, 2007). Additionally, sleep and circadian abnormalities are well documented (for review see Herzog-Krzywoszanska & Krzywoszanski, 2019; Morton, 2013; Y. Zhang et al., 2019). The disease is caused by a CAG triplet repeat expansion in the huntingtin gene, which is located on the short arm of chromosome four, and results in an expanded stretch of polyglutamine in the huntingtin protein (Pringsheim et al., 2012; Ross et al., 2014). The expanded huntingtin protein is neurotoxic and causes neuronal death, especially in the striatum and cortex (Blumenstock & Dudanova, 2020; F. O. Walker, 2007). The average number of CAG repeats in the general population is 16-20 (Pringsheim et al., 2012). Clinical manifestation of HD can occur in people with repeat lengths of 36 or greater with the age of onset inversely correlated with the length of the expansion (Pringsheim et al., 2012; Ross et al., 2014). Although the CAG repeat number is the primary determinant of age of onset, family members with identical repeat numbers can have significant variability in age of onset, meaning separate genetic and environmental factors can modify the age of onset of HD (Long et al., 2018; Trembath et al., 2010; Wexler et al., 2004). These genetic and environmental modifiers of disease onset are important targets for disease modification strategies (Long et al., 2018; Wexler et al., 2004) especially for people who may not benefit from current pharmacological therapeutic strategies (e.g., Rodrigues & Wild, 2020).

People with HD can experience cognitive and psychiatric symptoms, yet clinical diagnosis (manifest-HD) is based on the emergence of significant observable motor

dysfunction and confirmation of the presence of the expanded HD repeat via genetic testing (Cardoso, 2009; Pringsheim et al., 2012). Nonetheless, the cognitive symptoms in HD are a critical aspect of the disease because they can appear many years before clinical diagnosis (Maroof, Gross, & Brandt, 2011; J S Paulsen et al., 2008; Paulsen, Smith, & Long, 2013; Stout et al., 2014) and have a similar or larger effect on functioning and quality of life than motor symptoms (Eddy & Rickards, 2013; Ross, Pantelyat, Kogan, & Brandt, 2014; Van Liew, Gluhm, Goldstein, Cronan, & Corey-Bloom, 2013). Cognitive symptoms can also cause functional impairments prior to disease diagnosis (Duff et al., 2010; You et al., 2014). The period prior to clinical diagnosis will hereafter be referred to as premanifest-HD. Researchers have specifically focused on identifying genetic and environmental modifiers of cognitive, rather than just motor, symptom onset and expression because cognitive symptoms occur in premanifest-HD and affect functioning and quality of life (Long et al., 2018; Moss et al., 2017).

Attempts to identify modifiers of cognitive symptom onset have a key limitation, which is that the acquisition of the data needed for the studies takes several years because the studies require sample sizes in the thousands and because data collection relies on face-to-face assessments, limiting progress (Long et al., 2018; Moss et al., 2017). Fast, efficient, cost effective methods of phenotyping people with the expanded HD repeat are needed for large scale studies so that additional modifiers of cognitive aspects of disease onset can be identified and so responses to therapeutics can be efficiently monitored.

Remote, computerised, self-administered, performance-based cognitive tasks may provide a solution to some of the limitations inherent in the face-to-face cognitive assessments that are typically utilised in HD research (e.g., Long et al., 2018; Moss et al., 2017; Jane S Paulsen et al., 2014; Stout et al., 2012). Computerised cognitive tasks, including those on smart-phones and tablet devices, can enable rapid, inexpensive, accurate, and

convenient collection of cognitive data including the ability to conduct multiple assessments over time (Bauer et al., 2012; Dufau et al., 2011; Fredrickson et al., 2010; Koo & Vizer, 2019; Moore, Swendsen, & Depp, 2017; Zorluoglu, Kamasak, Tavacioglu, & Ozanar, 2015), which may allow for long term monitoring of response to therapeutic interventions. Additionally, modern smartphones can pair with relatively inexpensive actigraphy devices that track the physical activity and sleep habits of users (Ferguson, Rowlands, Olds, & Maher, 2015; Montgomery-Downs, Insana, & Bond, 2012; Takacs et al., 2014) and can also collect such data through their own sensors (Barnett et al., 2018; Tung et al., 2014). Consequently, researchers can pair ongoing monitoring of sleep and physical activity with ongoing assessment of cognitive functioning and capture dynamic interactions between these factors (Allard et al., 2014; Kaye et al., 2011; McCrae, Vatthauer, Dzierzewski, & Marsiske, 2012). Subsequently, researchers may then be able to determine the degree to which sleep and physical activity habits affect onset or progression of cognitive symptoms.

In my Introduction, I describe remote computerised assessment of cognition and explain that it can be used to collect reliable and valid measures of cognitive functioning data in a way that is more efficient and convenient than face-to-face assessment in the clinical setting. Additionally, I will highlight the importance of pairing cognitive assessment with measures of sleep and physical activity, which can readily be obtained remotely, due to their potential relationships with cognitive functioning. Utilisation of remote assessment to measure these factors is novel in HD; therefore, I also describe studies from other populations in which remote assessment has generated reliable, valid and informative data. Such studies have informed development of remote assessment methodology used in this HD project.

**1.1.1. Overview of the thesis aims and thesis structure.** The overarching goal of my thesis research was to design, create, test, and then utilise an exclusively remote means of collecting clinically and scientifically useful data on the day to day cognitive functioning, and

sleep and physical activity habits of people with premanifest- and early manifest-HD. The overall goal was broken down into specific aims. First, I aimed to demonstrate that people with the expanded HD repeat could complete the remote assessment protocol that we planned to use in the thesis research project (i.e., feasibility, acceptability, covered in Chapter Two). Next, I aimed to demonstrate that the cognitive tasks we developed for the research project would be reliable, and valid, across several forms of validity, and sensitive to cognitive functioning in HD. These aims are addressed in Chapters Three and Four. Finally, I also had the aim of adding to the limited literature on how sleep and physical activity habits relate to cognitive functioning in HD. This aim is addressed in Chapter Five.

I present my thesis in the following format:

Chapter One provides an overview of cognition in HD and what is known about the relationships between sleep-cognition and physical activity-cognition interactions in HD and why a better understanding of these interactions is needed. Chapter One also provides an overview of the current options that exist to conduct remote assessment of cognition, sleep, and physical activity, and explains that combined remote assessment of these factors in HD is needed but has not been attempted. Chapter Two describes the development and feasibility testing of the study protocol used in the thesis research. Chapters Three to Five are manuscripts reporting my findings relating to each of the main research aims. In Chapter Six, I integrate the results from Chapters Two to Five and discuss the broader clinical implications of the thesis research findings. Note, that this thesis is presented in line with Monash University guidelines and requirements as a ‘thesis by publication’. Some chapters comprise published papers, or papers under review. Due to the required format of the thesis, there will be some degree of unavoidable overlap and repetition of content for some chapters.

### 1.2. Cognition in Huntington's Disease

Severe cognitive decline is a primary symptom of Huntington's disease and causes marked functional deficits for people with HD, which can be separated from deficits stemming from the movement aspects of the disorder (Eddy & Rickards, 2015; F. O. Walker, 2007). The onset of cognitive decline is insidious and progressive and affects multiple domains of functioning, including, attention, processing speed, psychomotor speed, verbal and visual memory, decision making, and executive functioning (Ghilardi et al., 2008; Glikmann-Johnston, Fink, Deng, Torrest, & Stout, 2019; Ho et al., 2003; Stout et al., 2014; Stout, Rodawalt, & Siemers, 2001; Tabrizi et al., 2012; F. O. Walker, 2007). Cognitive deficits may be detected one to two decades before clinical diagnosis and as much as ten years before motor symptoms are detectable. The early onset of cognitive decline in people with the expanded HD repeat and the significant negative effects that those cognitive symptoms have on functioning and quality of life underscores the importance of cognitive assessment in HD as a means of understanding the natural course of the disease and in efforts to develop disease modifying treatments (Paulsen et al., 2008; Stout et al., 2012; Tabrizi et al., 2012).

The cognitive deficits that occur in HD are secondary to a range of neurochemical changes and to neuronal death, which is most severe in the striatum (Blumenstock & Dudanova, 2020; F. O. Walker, 2007). The neurochemical changes, striatal atrophy, and accompanying breakdown of fronto-striatal networks are associated with declines in a range of cognitive functions, including psychomotor functioning, processing speed, verbal learning, impulsivity, impaired judgement, planning, and rule learning (Bamford, Caine, Kido, Cox, & Shoulson, 1995; Blumenstock & Dudanova, 2020; Campodonico et al., 1998; Johnson, Potts, Sanchez-Ramos, & Cimino, 2017; Mörkl et al., 2016; O'Callaghan, Bertoux, & Hornberger, 2014; Peinemann et al., 2005; Starkstein et al., 1992).

Although an array of neurochemical changes and focal neuronal atrophy occur in the striatum, such changes are widespread and occur in other brain areas, such as the hippocampus and other cortical regions, with premotor and sensorimotor cortices particularly affected; neurochemical changes and atrophy in cortical regions are detectable prior to clinical diagnosis (Blumenstock & Dudanova, 2020; Glikmann-Johnston, Fink, et al., 2019; Rosas et al., 2005). As with striatal atrophy, the cortical atrophy that occurs in HD and the subsequent breakdown of cortical and cortical-subcortical networks throughout the brain are linked with a range of cognitive deficits (Glikmann-Johnston, Fink, et al., 2019; O'Callaghan et al., 2014; Rosas et al., 2005). For example, striatal and hippocampal atrophy and reductions in connectivity between the striatum and the hippocampus are suspected to contribute to spatial memory defects that occur in HD (Glikmann-Johnston, Fink, et al., 2019; Possin et al., 2017) and fronto-striatal disruptions may contribute to behavioural changes (O'Callaghan et al., 2014).

Regarding the trajectory of cognitive changes in HD, decline in the premanifest period may be rapid enough for clear changes to be measured over a period of as little as six months (Beste et al., 2013). More typically, a period of 24 months or greater may be needed for clear declines in cognitive performance to be detectable (Ho et al., 2003; Stout et al., 2012). Because noticeable cognitive declines, at least in the premanifest stage, can take 12 – 24 months to become apparent, studies assessing cognitive functioning and cognitive symptom progression commonly use a protocol of once-yearly assessments (e.g., Beste et al., 2013; Ho et al., 2003; Stout et al., 2012; Stout et al., 2014; Tabrizi et al., 2013). Such studies have informed about cognitive symptom onset and progression, but are limited by the lengthy gap in between assessments and by the fact that all assessments were conducted in the clinic. Once yearly in-clinic assessments provide little information about the day-to-day cognitive and functional independence of a person in the pre-manifest period. As a result, these

longitudinal studies may lack ecological relevance. The use of infrequent assessments also limits the understanding of how cognitive function may vary across a day or week, in response to environmental and contextual factors such as mood, physical activity levels, or a poor night's sleep (Allard et al., 2014; Kaye et al., 2011; McCrae et al., 2012). Such information would be very useful for developing non-pharmacological means of managing cognitive symptoms. For example, if a variable sleep schedule negatively affects cognition in HD, as it can in the general population (Bei, Wiley, Trinder, & Manber, 2016), then clinicians could promote a regular sleep schedule as a way of maximising the cognitive functioning of their patients. Altogether, pairing frequent in-home assessment of cognition with ongoing monitoring of sleep and physical activity habits has the potential to provide a rich understanding of how sleep and physical activity habits and day to day cognitive functioning interact in HD. Moreover, such an approach may capture how cognitive functioning fluctuates over short periods, even with time of day, and how people with the expanded HD repeat are functioning in the community, rather than in the clinic.

Smartphone-based cognitive assessment has the potential to provide the means to conduct frequent and convenient in-home assessment because computerised cognitive assessment can enable rapid, inexpensive, accurate, and convenient means of collecting cognitive data (Bauer et al., 2012; Dufau et al., 2011; Fredrickson et al., 2010; Moore et al., 2017; Zorluoglu et al., 2015). Additionally, remote cognitive assessment can be paired with remote assessment of sleep and physical activity habits. As I will describe below, sleep and physical activity habits can have important implications for cognitive functioning, both in the general population and in Huntington's disease. Assessment of these factors and how they relate to cognitive functioning will be important for efforts to understand long-term environmental modifiers of cognition in HD but may also have important implications for how people with the expanded HD repeat can manage their day-to-day cognitive symptoms.

**1.2.1. Sleep disturbance and cognition in Huntington's disease.** Sleep disturbance is common in HD (Arnulf et al., 2008; Herzog-Krzywoszanska & Krzywoszanski, 2019; Y. Zhang et al., 2019). Disturbed sleep and changes in sleep wake cycles have been documented with self-report measures (Goodman, Morton, & Barker, 2010; Taylor & Bramble, 1997), actigraphy (Aziz, Anguelova, Marinus, Lammers, & Roos, 2010; Diago et al., 2018; Goodman et al., 2011; Morton et al., 2005), and polysomnography based methodologies (Arnulf et al., 2008; Cuturic, Abramson, Vallini, Frank, & Shamsnia, 2009; Wiegand, Möller, Lauer, et al., 1991; Wiegand, Möller, Schreiber, Lauer, & Krieg, 1991; Y. Zhang et al., 2019). The amount of sleep disturbance experienced by people with the expanded HD repeat can vary by disease stage. Those in the manifest stage often experience a greater degree of sleep disturbance than those in the premanifest stage who may have greater sleep disturbance than healthy controls and non-affected family members (Aziz et al., 2010; Goodman & Barker, 2010; Herzog-Krzywoszanska & Krzywoszanski, 2019; Lazar et al., 2015).

Although, sleep disturbance in manifest-HD can be worse than in premanifest-HD, there appears to be no homogenous pattern of sleep disorders in HD (Happe & Trenkwalder, 2002; Herzog-Krzywoszanska & Krzywoszanski, 2019; Morton, 2013) making a description of the course of sleep disturbances difficult. For example, Goodman et al. (2010) classified 35.3% and 26.5% of 66 study participants, with varying stages of HD, as having mild or severe sleep disturbance, respectively, and as being more sleep disordered than their carers or healthy control participants. Somewhat differently, Taylor and Bramble (1997) reported that 87.8% of respondents to a mail survey reported experiencing sleep problems. One cannot extract a clear picture of the type and severity of sleep disturbance across the two studies because neither study reported the proportion of premanifest- versus manifest-HD participants that were involved in their studies. Moreover, methodologies between the two studies differed, making it impossible to compare results directly. Goodman et al. (2010) used

a scoring system to rate the degree of sleep disturbance in their participants, whereas, Taylor and Bramble (1997) reported on the prevalence of any type of sleep problem reported by participants in their study. Additionally, disordered REM sleep has been observed in some studies (Arnulf et al., 2008; Hansotia, Wall, & Berendes, 1985; Silvestri et al., 1995) but also as absent (Wiegand, Möller, Lauer, et al., 1991). Nonetheless, changes in circadian rhythms, particularly a tendency towards waking later in the morning, appears to be a common early symptom (Aziz et al., 2010; Diago et al., 2018; Herzog-Krzywoszanska & Krzywoszanski, 2019). The lack of a homogenous account of sleep disturbance in HD may be due to studies including participants of varying disease progression, utilisation of different study procedures and methodologies, but also the small sample sizes used in many studies.

Large prospective studies that collect ongoing information about how people sleep in their home environment may help to provide a clearer picture of the natural course of sleep changes in HD, especially as wearable sensors now exist that can measure electroencephalogram and electrocardiogram signals while worn in the home (Arnal et al., 2019; Boe et al., 2019). Additionally, prospective home-based monitoring of sleep provides an ecologically valid understanding of sleep habits and how they relate to cognition (McCrae et al., 2012). McCrae et al. (2012) argue that sleep diaries have an important role to play in understanding sleep habits of people with sleep complaints. In contrast, Y. Zhang et al. (2019) argue that comprehensive polysomnography studies are needed to investigate and understand the sleep changes in HD. A combined approach is probably the best practice as objective and subjective measures of sleep can relate to non-overlapping aspects of sleep functioning (L. Zhang & Zhao, 2007) and hence the inclusion of both methods of assessment should give the most thorough assessment of sleep in HD. Importantly, subjective and objective measures of sleep can be combined in studies which occur outside of the laboratory since sleep diaries and sleep questionnaires can easily be completed at home (Aziz et al.,

2010; Taylor & Bramble, 1997; L. Zhang & Zhao, 2007) and because several wearable sensor devices are available which can measure a range of sleep variables from sleep timing through to measurement of electrooculography and electrocardiogram signals (Arnal et al., 2019; Boe et al., 2019; Ferguson et al., 2015; Maskevich, Jumabhoy, Dao, Stout, & Drummond, 2017). If these technologies can be combined with cognitive assessment that can also be completed in the home, a detailed understanding of the sleep habits of people with the expanded HD repeat can be established, as could an understanding of how sleep habits relate to cognitive functioning in the home.

The presence of disturbed sleep in HD has important implications for cognitive functioning for several reasons; a) disordered sleep is associated with negative cognitive outcomes across the human lifespan (Antoniadis, Ko, Ralph, & McDonald, 2000; Beebe, 2011; Ellenbogen, 2005; Killgore, 2010; Yaffe, Falvey, & Hoang, 2014); b) sleep disturbances are expected to exacerbate the neurological symptoms of HD (Goodman & Barker, 2010); c) later habitual wake times in HD samples correlate with poorer cognitive functioning (Aziz et al., 2010; Diago et al., 2018), and; d) disturbed sleep may be associated with reduced thalamic brain volumes in manifest-HD (Baker et al., 2016). To address these issues, characterisation is needed of both the nature of sleep disturbance in HD and any associated effects on symptomatology and functioning. Of particular importance will be understanding how sleep disturbance relates to cognitive functioning in HD, which is currently understudied.

Several studies have investigated sleep disturbance in HD, but few have investigated the link between sleep disturbance and cognitive functioning. Two HD studies have investigated links between sleep and cognition in humans (Aziz et al., 2010; Diago et al., 2018) and additional studies have been conducted using mouse models (Pallier et al., 2007; Pallier & Morton, 2009). Regarding the relationship between sleep and cognition in humans,

similar studies by Aziz et al. (2010) and Diago et al. (2018) observed a relationship between later habitual wake time and poorer cognitive functioning. Aziz and colleagues mailed a series of sleep questionnaires to participants, which included questions about their sleep quality in the past month and their usual bed and wake times and then conducted an in-clinic assessment three weeks later including cognitive tasks from the Unified Huntington's Disease Rating Scale (UHDRS, Huntington Study Group, 1996). Later habitual wake up time associated with poorer cognitive performance on several tests, including tests of processing speed and higher-level attention and inhibition. Diago et al. (2018) utilised a very similar approach to assessing sleep-cognition interactions. Participants ($n$ = 23 premanifest-HD and $n$ = 15 manifest-HD) completed sleep questionnaires and UHDRS cognitive tasks, and also observed associations between later habitual wake time and cognitive performance on tests of processing speed and inhibition. Importantly, Diago et al. (2018) included premanifest-HD participants in their correlations, whereas Aziz et al. (2010) only included manifest-HD participants, meaning that the combined findings indicate that the relationship between habitual wake time and cognition appears to be important for HD across a range of the disease severity spectrum.

HD mouse models provide further evidence that sleep habits, and in particular sleep schedules, have important relations to cognitive functioning in HD (Pallier et al., 2007; Pallier & Morton, 2009) similar to what has been found in the human studies by Aziz et al. (2010) and Diago et al. (2018). For example, Pallier and Morton (2009) reported improvements in cognition and apathy of R6/2 HD mice following pharmacological treatment to correct disintegrated sleep/wake patterns. Pallier and Morton controlled the sleep patterns of the mice by treating them with Alprazolam, to put them to sleep, and Modafinil, to wake them. They assessed cognition using a two-choice swim tank task, with an acquisition and a reversal phase, and assessed apathy by testing if the mice would wake in response to

mild cage disturbance. The mice with the amended sleep schedule outperformed their untreated cage mates on the two-choice swim task, indicating they had better spatial memory functioning. Further, the treated mice, which were previously resistant to rousing (Pallier et al., 2007), had identical rousing behaviour to wild-type mice following pharmacological treatment, indicating apathy had been reduced.

The above studies represent most of what is known about sleep-cognition relationships in HD and as such, sleep-cognition relationships are understudied and require further investigation. Nonetheless, the broader sleep literature clearly demonstrates that sleep disturbances have negative consequences for cognitive functioning, even in healthy populations. For example, sleep disturbance compromises memory and learning (Born & Wagner, 2004; Diekelmann & Born, 2010; M. P. Walker & Stickgold, 2014), vigilance and alertness (Jewett, Dijk, Kronauer, & Dinges, 1999; J. Lim & Dinges, 2008), emotional functioning (Kahn-Greene, Lipizzi, Conrad, Kamimori, & Killgore, 2006; Killgore, Balkin, & Wesensten, 2006), and working memory (J. Lim & Dinges, 2010). Therefore, sleep disturbance can logically be expected to negatively affect cognition in HD. Thus, hypotheses that sleep disturbance in HD will exacerbate symptomatology (Goodman & Barker, 2010; Y. Zhang et al., 2019) seem reasonable as do suggestions that treatments of sleep disturbance may have an ameliorating effect on symptoms (Morton, 2013; Pallier & Morton, 2009; Videnovic, Leurgans, Fan, Jaglin, & Shannon, 2009). Better characterisation of the relationship between sleep and cognition is needed in HD to define which aspects of sleep or sleep disturbance are most closely related to cognitive function in HD, and to inform treatment strategies. For instance, habitual wake time appears to be associated with cognitive functioning (Aziz et al., 2010; Diago et al., 2018), and may be a target for interventions aimed at improving cognitive functioning in HD. Large scale and prospective remote

assessment of sleep and cognition can provide a means of disentangling of these relationships.

**1.2.2. Physical activity and cognition in Huntington's disease.** Insufficient physical activity associates with poorer cognitive functioning across the human lifespan and is a risk factor for the development of dementias including Alzheimer's disease and Parkinson's disease (Colcombe & Kramer, 2003; Esteban-Cornejo, Tejero-Gonzalez, Sallis, & Veiga, 2015; Hamer & Chida, 2009; Lautenschlager, Cox, & Kurz, 2010; Sibley & Etnier, 2003). In particular, physical exercise can benefit cognition and protect against neurodegeneration by increasing neuroplasticity, neuroprotection, and neurorestorative processes (Hirsch, Iyer, & Sanjak, 2016; Hötting & Röder, 2013). Despite these known benefits, the relationship between physical activity and cognition in HD is understudied. One observational study in HD has investigated the relationship between physical activity and cognition and indicated that higher amounts of physical activity were related to better cognitive functioning (Wallace et al., 2016). Several physical activity interventions have been conducted in HD, but few of these had cognitive tests as primary outcome measures and the results in terms of improvements in cognition have been mixed (for review, see Fritz et al., 2017). Additionally, lifestyle passivity, meaning low level engagement in physically and cognitively demanding activities, may be associated with earlier onset of motor symptoms (Trembath et al., 2010). Mouse studies also provide some evidence that exercise can benefit cognition in HD. Mice raised in enriched environments, which includes areas to play, running wheels and other toys, had better spatial memory than mice kept in enclosures (Nithianantharajah, Barkus, Murphy, & Hannan, 2008; Wood et al., 2010). The effect of increased exercise on cognition cannot be disentangled from the effect of the enriched environment, thus these studies do not definitively show that exercise alone leads to improvements in spatial memory.

As mentioned, physical activity intervention studies in HD that measured cognition as an outcome variable have had mixed results in terms of whether or not the interventions resulted in improvements to cognitive functioning. Only one study of six studies reviewed by Fritz et al. (2017) had cognition as a primary outcome measure. Cruickshank and colleagues (2015) conducted a physical activity intervention that resulted in some cognitive improvements. The intervention ran for nine-months and with 15 participants with manifest-HD. The intervention included one supervised and three self-directed exercise sessions per week, along with an occupational therapy session (Cruickshank et al., 2015). The authors reported that following the intervention, study participants had improved performance on the Hopkins Verbal Learning Test (HVLT), a measure of verbal learning and memory (Brandt & Benedict, 2001), and increased grey matter volume in the right caudate and bilaterally in the dorsolateral prefrontal cortex. The participants did not improve on other cognitive tests, such as the Trail Making Test and Symbol Digits Modalities Test, and no control cohort was available for comparison. Additionally, no significant improvements in cognition were noted following a similar intervention by the same group (Thompson et al., 2013) and no cognitive improvements were noted following three other physical activity intervention studies (Busse et al., 2013; Quinn et al., 2014; Quinn et al., 2016). Taken together, the findings from the above studies seem to indicate that physical activity may not benefit cognition in HD as it does in other populations. It is important to remember that thus far there have been relatively few physical activity intervention studies, with fewer still focusing on cognitive functioning outcomes. Moreover, available studies have often been limited by small sample sizes, lack of control groups, and variations in intervention approach and intervention outcome measurement (Fritz et al., 2017). Nonetheless, there are several reasons to continue investigating physical activity-cognition interactions in HD and to expect that physical activity may benefit cognition, particularly since studies in other neurodegenerative disease

indicate that physical activity does benefit cognition. A large body of evidence demonstrates that lifetime physical activity is associated with reduced risk, delayed onset, and attenuated progression of neurodegenerative diseases other than HD, such as Alzheimer's disease and Parkinson's disease (Brown, Peiffer, & Martins, 2013; Hamer & Chida, 2009; Paillard, Rolland, & de Souto Barreto, 2015; Rovio et al., 2005). Evidence also shows that physical activity can have direct benefits for cognition for healthy older adults (Colcombe & Kramer, 2003; Kramer et al., 1999) and people with dementia and other neurodegenerative disease (Heyn, Abreu, & Ottenbacher, 2004; Lautenschlager et al., 2010; Vidoni & Burns, 2015) and across the lifespan in the general population (Colcombe & Kramer, 2003; Esteban-Cornejo et al., 2015; Sibley & Etnier, 2003). Considering some specific examples, amongst a large Finnish cohort of 1449 persons aged 65 – 79 years, engaging in leisure time physical activity at least twice a week during mid-life was associated with a significantly reduced risk of dementia (odds ratio [OR] 0.48 [95% CI 0.25–0.91] and Alzheimer's disease [OR] 0.38 [0.17–0.85] (Rovio et al., 2005). The findings of this study were very robust because Rovio and colleagues controlled for age, sex, education, follow-up time, locomotor disorders, apolipoprotein E genotype, vascular disorders, smoking, and alcohol use. Additionally, the study was effectively prospective because the researchers had access to participants' self-reported physical activity data as part of their participation in an earlier study, improving reliability of data over a retrospective recall methodology. Furthermore, in a meta-analysis of 16 prospective studies investigating associations between physical activity and onset of Alzheimer's and Parkinson's disease, overall physical activity levels were inversely associated with risk of both Alzheimer's and Parkinson's disease (Hamer & Chida, 2009). Specifically, relative risk of Alzheimer's disease for the highest physical activity category compared to the lowest was 0.72 [95% confidence interval (CI) = 0.60–0.86, $p < 0.001$] and the relative risk of Parkinson's disease was 0.82 [95% CI = 0.57–1.18, $p = 0.28$]. Hence, the

association between physical activity and reduced risk of developing dementia, Alzheimer's disease, and Parkinson's disease appears robust. Exercise may protect against neurodegenerative processes in disorders such as Parkinson's disease and Alzheimer's by enhancing neurogenesis and synaptogenesis (Paillard et al., 2015). Given that exercise seems protective of the onset of Parkinson's and Alzheimer's diseases, it is reasonable to expect that exercise may also have a protective role in HD. Indeed, Trembath et al. (2010) demonstrated that lifestyle passivity (low level engagement in physically and cognitively demanding activities) associated with earlier onset of motor symptoms. The possibility that cognitive symptom onset will be affected in a similar way seems high but needs to be properly investigated.

Exercise may be a protective factor for HD in the long term, but it may also improve cognition in the short term, as has been observed in other populations (Colcombe & Kramer, 2003; Heyn et al., 2004; Kramer et al., 1999). As an example, Colcombe and Kramer (2003) analysed 18 prospective studies which assessed the effect of structured cardiovascular training interventions on cognitive performance in healthy older adults, aged 55 and above. Results indicated that cognitive function of exercisers improved significantly more than control participants for executive functioning, speeded tasks, and visuospatial tasks. Taken together, exercise likely plays an important role in delaying onset of neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease and may also confer benefits on cognition in ageing populations following intervention. Potentially then, both of these relationships may hold in Huntington's disease and require investigation. As with sleep, remote assessment of physical activity will allow collection of large amounts of data which can be used to determine the relationship between physical activity and cognition in HD. Wearable actigraphy devices provide a viable means of remotely assessing physical activity (Montgomery-Downs et al., 2012; Takacs et al., 2014; Wallace et al., 2016) as is explained in

detail below. The importance of understanding any benefit that physical activity may have on delaying onset of cognitive symptoms in HD is underscored by remembering that the cognitive symptoms in HD lead to functional declines prior to diagnosis (Duff et al., 2010; You et al., 2014) and that they cause significant reductions in quality of life and functioning (Eddy & Rickards, 2013, 2015; Ross et al., 2014; Van Liew et al., 2013). Therefore, long-term prospective studies are needed to determine associations between physical activity, or other lifestyle factors, and onset of cognitive symptoms. Meanwhile, shorter term observational studies can be used to probe contemporaneous associations between physical activity and cognition by combining prospective assessment of cognition and sleep over the period of one or several weeks, with ongoing assessment of cognition. As will be described next, these goals can be achieved by the use of mobile phone and wearable sensor technologies and will allow ongoing assessment to be achieved without requiring participants to attend the laboratory or clinic.

### 1.3. Remote and computerised Assessment of Cognition, Sleep, and Physical Activity

**1.3.1. Computerised assessment of cognition.** The most common method of measuring cognitive performance in neuropsychological assessment is the use of pen and paper-based assessments (Parsey & Schmitter-Edgecombe, 2013). Due to this, automated or computerised cognitive assessment methods are often viewed as experimental, or novel in nature, which serves to undermine their usage (Wesnes, 2014). The use of automated testing procedures, however, predates many of the tests used in clinical neuropsychology today (Wesnes, 2014). For example, Wesnes (2014), reports that in 1868, Francois Donders developed a complex mechanical device which was used to accurately measure verbal or manual response times of human participants following presentation of a stimulus, such as a coloured light, letter-symbol, or sound. Additionally, psychologists have been using computerised cognitive tasks since as early as the 1960s, and more extensively since the

development of the personal computer in the 1980s (Hagelkruys et al., 2016; Wesnes, 2014), including extensively in evaluation of cognitive functioning in military and sport psychology settings (Parsey & Schmitter-Edgecombe, 2013).

Currently, computerised neuropsychological testing is used with a variety of populations, ranging from children with ADHD (Chamberlain et al., 2011), to assessing cognitive function in older adults (Darby et al., 2014; Eve & De Jager, 2014; Y. Y. Lim et al., 2012; Tierney et al., 2014; Tornatore, Hill, Laboff, & McGann, 2005) and for measurement of cognitive impairment in samples diagnosed with psychiatric illnesses (Hamo, Abramovitch, & Zohar, 2018; Iverson, Brooks, Langenecker, & Young, 2011; Sweeney, Kmiec, & Kupfer, 2000). Computerised tests are also used to evaluate the effects of sports-related concussion (Collie, Makdissi, Maruff, Bennell, & McCrory, 2006; Makdissi et al., 2001), and for cognitive assessment of people with dyslexia (Hagelkruys et al., 2016). Thus, because automated and computer based psychological testing procedures have been utilised for over 100 years, and 50 years, respectively, rejection of their utilisation based on concerns that they are novel or overly experimental is unjustified. Moreover, as noted by Wesnes (2014) and Zygouris and Tsolaki (2015), provided that a measure has satisfactory psychometrics, the decision to use a computerised or pen and paper based test, or, by extension, a combination of the two, should be driven by selection of the test which is best suited to measure change over time, as well as selecting the measure that can best be used to achieve the aim(s) of the clinician or researcher. For instance, remote computerised assessment may be best suited to continued measurement of cognitive functioning in HD, where mobility and the ability to attend in person testing can become steadily compromised over time due to loss of driving ability (Beglinger et al., 2010; Farrell et al., 2019; Vaccarino et al., 2011).

The wide usage of computerised neuropsychological measures may be in part due to the many potential benefits that they possess in comparison to pen and paper-based tests, such as the symbol digit modalities and Stroop interference tests which are included in the UHDRS cognitive battery. These properties mean that computerised cognitive assessment can potentially be used to measure and monitor cognitive functioning in large neuro-epidemiological studies (Fredrickson et al., 2010). In a joint position paper on the usage of computerised neuropsychological measures (Bauer et al., 2012), the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology in the US identified six strengths of computerised neuropsychological tests. These strengths included: a) the capacity to test a large number of individuals quickly; b) ready availability of assessment services without advanced notice; c) the ability to precisely measure performance on time-sensitive tasks, such as reaction time; d) potentially reduced assessment times through the use of adaptive testing protocols; g) automated data exporting for research purposes, h) increased accessibility to patients in areas or settings in which professional neuropsychological services are scarce; and i) the ability to integrate and automate interpretive algorithms such as a decision rule for determining impairment or statistically reliable change. Thus, computerised cognitive testing is well suited to efficiently collect large amounts of detailed data related to cognitive functioning, even in a geographically dispersed sample.

The use of algorithms can reduce floor and ceiling affects and aid differential diagnosis by being programmed to detect impairment patterns indicative of specific disorders (Bauer et al., 2012; Mohr, Walker, Randolph, Sampson, & Mendis, 1996; Parsey & Schmitter-Edgecombe, 2013; Wild, Howieson, Webbe, Seelye, & Kaye, 2008). Additionally, the reduction of floor effects means computer-based tests may be more suitable for detecting mild cognitive impairment and early signs of dementia, than pen and paper-based tests (Wild

et al., 2008). This makes computerised cognitive testing ideal for detecting and measuring onset and progression of cognitive symptoms in premanifest-HD, as well as responses to interventions, because impairments, and any alterations in cognitive functioning, may initially be very mild, and because the level of difficulty of the cognitive tests can be adjusted to suit the disease stage of patients and participants.

A common caveat surrounding the use of computerised testing is that many newly developed computerised tests should be used with caution due to some of them being insufficiently normed or validated (Bauer et al., 2012; Hagelkruys et al., 2016; Schlegel & Gilliland, 2007; Wild et al., 2008). This is clearly not a criticism that is specific to computerised tests, but rather it relates to the development of all psychological testing batteries. Another common criticism of computerised tests is that they may be intimidating and unfamiliar to some segments of the populations, especially older adults (Wesnes, 2014). However, a variety of tests batteries designed to measure dementia and cognitive impairment have been successfully used with older-aged participants, and by participants with cognitive impairment (Darby et al., 2014; Eve & De Jager, 2014; Fredrickson et al., 2010; Koo & Vizer, 2019; Y. Y. Lim et al., 2012; Mohr et al., 1996; Stout et al., 2014; Tierney et al., 2014; Zorluoglu et al., 2015; Zygouris & Tsolaki, 2015). Therefore, computerised cognitive batteries clearly are suitable for use with neurodegenerative samples and with people who may have limited experience with computer technology. For example, Collerton et al. (2007) randomised elderly participants ($N$ = 81, all aged 85) to either a computerised cognitive battery or a pen and paper battery. Those that completed the computerised battery rated it as less difficult and more acceptable than those that completed the pen and paper battery. Moreover, a greater percentage of participants completed the computerised battery than completed the pen and paper battery (100% v 91%) and the experimenters rated those completing the computerised battery as less likely to appear distressed compared to those

completing pen and paper tests. Importantly, the scores on both batteries correlated significantly, medium to strong effect sizes, with scores on the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) and participants reporting subjective memory problems scored lower on all tasks, with the difference reaching statistical significance on two of the computerised tasks. Thus, the results indicate that cognitive data can be collected via computer in a way that is at least as valid and acceptable to older participants as data collected on pen and paper tasks. Together with many of the studies mentioned above, this demonstrates that computerised batteries can be used in many patient groups, including those with cognitive impairments.

One final advantage of computerised cognitive tasks is the ability to mathematically model data (e.g., data from binary decision making tasks). These models are often more sensitive to individual cognitive change than paper-based tools used in clinic settings (Busemeyer & Diederich, 2010; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016; Voss, Nagler, & Lerche, 2013) and may be particularly useful for detecting the mild cognitive changes that occur in premanifest-HD. Modelling cognitive data with drift diffusion methods (Ratcliff, 1978; Ratcliff et al., 2016) produces variables that measure a person's speed of information encoding and the amount of information they require before making a decision (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Voss et al., 2013). Importantly for the study of cognition in HD, the striatum is thought to modulate the amount of information a person choses to consider prior to making a decision (Bogacz et al., 2010; Frank, 2006). For example, decision making cautiousness or rashness can be manipulated via deep brain stimulation or by changing doses of dopaminergic medication in Parkinson's disease patients (Frank, Samanta, Moustafa, & Sherman, 2007; Herz et al., 2018; Huang et al., 2015). Severe striatal atrophy and behavioural changes, such as impulsivity, occur in HD (Stout et al., 2001; F. O. Walker, 2007), and so cognitive modeling of decision

making tasks from HD patients, may help to determine if striatal and behavioural changes in HD are linked as is suspected (O'Callaghan et al., 2014).

**1.3.2. Remote computerised assessment of cognition.** To date, the majority of studies utilising computerised cognitive assessment were administered face-to-face in the clinic. More recently, several mobile phone and tablet-based assessment tools have been developed that can be used by research participants in their own homes (Allard et al., 2014; Darby et al., 2014; Eve & De Jager, 2014; Jongstra et al., 2017a; Lipsmeier et al., 2018; Rentz et al., 2016; Zygouris et al., 2017). Shifting computerised assessment to be self-administered by people in their own homes utilises two of the greatest strengths of computerised assessment, including increasing the capacity to test a large number of people quickly, and improved accessibility to patients in areas or settings in which professional neuropsychological services are limited (Bauer et al., 2012). These qualities are apt for responding to the urgent need for simple, inexpensive methods of detecting early neurodegenerative decline (Bauer et al., 2012; Sperling et al., 2011). Also, mobile platforms enable repeated assessment of cognition, which can provide a more reliable description of cognitive difficulties than is possible with infrequent assessments in the clinic (Allard et al., 2014). Remote computerised assessment is a relatively new direction in the field of computerised cognitive assessment. Nonetheless, studies utilising unsupervised cognitive assessment protocols have been published with promising result and some with important implications for the idea of conducting remote cognitive assessment in HD. I briefly review some key remote assessment studies next.

Eve and De Jager (2014) developed and assessed a laptop-based, self-administered, online cognitive screening tool, the Cognitive Function Test (CFT), in a two-stage process. The first stage involved 50 healthy adults aged 50 – 65 years, with no dementia diagnosis or significant memory complaints. Participants completed the CFT and pen and paper tests of cognition in their own homes under the supervision of an administrator. In the second stage

of the study, a separate cohort of 195 adults, aged, 50 - 65 years, completed the CFT online, with no supervision or prior training. Scores from these two cohorts were compared. Eve and De Jager (2014) reported two key results. First, total scores on the CFT were strongly and positively correlated with total scores from the pen and paper tests, indicating that the CFT shared concurrent validity with the pen and paper tests. Second, scores obtained on the CFT by the supervised and unsupervised samples were similar, indicating that test performance did not vary based on the presence or absence of a test administrator. The key finding from this study, as it relates to the prospect of remote assessment in HD, is that cognitively healthy participants completed the cognitive battery in their own home without prior training or supervision and without a loss of validity in results. This indicates that the same approach may be possible with a HD sample. An important caveat, however, is that participants in Eve and De Jager's study were cognitively healthy, and so it is unclear if a cognitively impaired sample would have the same success.

Darby et al. (2014) conducted a similar study, testing an online cognitive battery, the CogState Brief Battery (CBB) with 150 participants aged between 55 – 83 years. The participants were experienced with the cognitive battery, having previously completed the battery under supervised conditions as part of an earlier study by the same research group (Fredrickson et al., 2010). In the Darby et al. (2014) study, participants completed the battery, unsupervised, once a month on a home computer for a period of twelve months. Because the psychometric properties of the CBB had been previously established (Fredrickson et al., 2010), the primary aim of Darby and colleagues was to assess the feasibility of asking the participants to conduct repeated self-assessment sessions over the 12 month period of the study. The eventual attrition rate for the study was moderately high. Only 63% of the 150 participants continued to complete testing at three months, 58% at six months, and less than 43% of the 150 participants completed testing at the 12-month period. The key finding from

this study was that, despite a high attrition rate, Darby et al. (2014) showed that an internet-based online assessment could be successfully implemented. Their attrition rate highlights the need to develop methods for keeping participants engaged if planning to conduct longitudinal assessment. Similar to Eve and De Jager (2014), the sample was cognitively healthy and so it remains difficult to draw a conclusion about the feasibility of attempting a similar approach with a cognitively impaired or neurodegenerative population. In another more recent study, Lipsmeier et al. (2018) demonstrated that people with a neurodegenerative disease can complete unsupervised home-based assessment. Lipsmeier et al. (2018) deployed a smartphone-based motor symptom assessment tool in 44 people with Parkinson's disease. The application contained six tests which the participants were asked to complete on a daily basis for six months. At the start of the study, Lipsmeier and colleagues trained the participants to use the application. Similar to Darby et al. (2014), Lipsmeier and colleagues assessed the feasibility of asking the study participants to self-administer tests at home without any supervision. Overall, 61% of all possible tests were completed over the six-month study period, and 90% of participants completed a test at least once every four days. The results from Lipsmeier and colleagues' study demonstrate the feasibility of remote, smartphone-based assessment in a neurodegenerative group, including task self-administration without supervision. Thus, Lipsmeier and colleagues' results suggest that, *with prior training*, remote cognitive assessment is likely to be feasible, particularly for people in the premanifest or early manifest stage of HD. The feasibility of attempting remote cognitive assessment in HD *without* providing any formal training in the clinic remains unknown. If remote cognitive assessments can be completed without first training patients or participants, such an approach would maximise one of the main benefits of remote assessment, which is reduced attendance at hospitals by patients and participants (Klimova, 2017; Koo & Vizer, 2019). Altogether, untrained remote cognitive assessment has several potential benefits for

the field of HD, but the feasibility of the approach must be demonstrated before it is attempted on a large scale.

**1.3.3. Remote assessment of sleep and physical activity.** Coupling remote assessment of cognition with assessment of lifestyle and behavioural factors such as sleep and physical activity will allow investigation of the relationships between these factors. Additionally, in the longer term, the large amounts of data that can be rapidly and conveniently collected in this way will be beneficial for attempts to uncover environmental modifiers of cognitive age of onset in HD. A variety of means of assessing environmental and lifestyle factors remotely currently exist (Arnal et al., 2019; Barnett et al., 2018; Boe et al., 2019; Boulos, Wheeler, Tavares, & Jones, 2011; Estudillo-Valderrama, Roa, Reina-Tosina, & Naranjo-Hernández, 2009; Kaye et al., 2011; Worringham, Rojek, & Stewart, 2011).

A small number of studies have attempted remote monitoring of sleep and/or physical activity in HD samples using actigraphy devices (Goodman et al., 2011; Morton et al., 2005; Wallace et al., 2016). Actigraphy is a form of monitoring sleep and activity that uses a portable, usually wrist mounted, device that can be conveniently worn for long periods of time (Ancoli-Israel et al., 2003; Chesson Jr, Coleman, Lee-Chiong, & Pancer, 2007; Sadeh, 2011). Actigraph devices use accelerometers to monitor movement, and activity data is converted into sleep/wake scores using computerised scoring algorithms (Ancoli-Israel et al., 2003; Sadeh, 2011). Movement can be sampled as much as several times per second and devices may have enough memory to remotely record activity for several weeks (Ancoli-Israel et al., 2003). Actigraphy is accepted to provide sensitive measures of sleep/wake patterns and sleep disturbance in various sleep disorders as well as with medical and neurobehavioural disorders, especially in situations where polysomnography, the gold standard method of measuring sleep is not available or appropriate (Chesson Jr et al., 2007;

Ferguson et al., 2015; Sadeh, 2011). In HD studies Actigraphy should be used to measure sleep wake total sleep time and sleep efficiency, because night time chorea can confound the measurement of total sleep time and sleep efficiency (Maskevich et al., 2017; Townhill et al., 2016). In summary, actigraphy devices are useful and valid tools for conducting ongoing measurement of sleep and activity data, and previous studies have shown that people with the expanded HD repeat can operate these devices properly over the course of research studies. Hence, it is feasible to couple remote assessment of sleep and cognition with remote assessment of cognition in HD. Moreover, coupling prospective measurement of sleep and physical activity, from actigraphy and self-report measures, with frequent cognitive assessment will provide a detailed account of dynamic interactions between cognition, sleep, and physical activity, both in short- and long-term studies (Allard et al., 2014; Kaye et al., 2011; McCrae et al., 2012), which can then inform about short- and long-term modifiers of cognitive symptoms in HD.

### 1.4. Conclusion and Summary of Research Aims

Overall, my thesis introduction has explained that cognitive symptoms are a primary and debilitating feature of HD, but that current efforts to identify modifiers of cognitive symptom progression and onset are limited by utilising infrequent clinic-based assessments that require several years to accumulate useful datasets. Moreover, infrequent assessment of cognition, sleep, and physical activity in HD mean that understanding how these variables fluctuate in the day to day lives of people with the HD CAG expansion is limited. I have argued that these limitations can be addressed by combining remote computerised assessment of cognition with remote assessment of sleep and physical activity. Lastly, I have explained that at present no one has attempted to combine remote assessment of cognition, sleep, and physical activity in HD. This last point leads into the aims of the thesis research, which were presented above and are again summarised here.

First, I aimed to demonstrate that people with the expanded HD repeat could complete the protocol (i.e., feasibility, acceptability, covered in Chapter Two). Next, I aimed to demonstrate that the cognitive tasks we developed for the research project would be reliable, and valid, across several forms of validity, and sensitive to cognitive impairment in HD compared to controls. These aims are addressed in Chapters Three and Four. Finally. I also had the aim of adding to the limited literature on how sleep and physical activity habits relate to cognitive functioning in HD. This aim is addressed in Chapter Five.

# Chapter Two: Study One

## Development and Initial Feasibility Testing of the Remote Assessment Protocol

### 2.1. Explanatory Notes

In this chapter I present an account of the development and initial feasibility testing of the remote assessment protocol that was used in the thesis research. In Chapter One, I explained that, to my knowledge, no one has attempted to combine remote assessment of cognition, sleep, and physical activity in HD and that, to my knowledge, no one has attempted remote cognitive assessment in a cognitively compromised sample without providing any in-person training. Because of this, we took a careful and considered approach to the development of our remote assessment protocol to make sure that people with the expanded HD repeat would find the protocol acceptable. This chapter provides an overview of that developmental work, which we hope will be a useful guide to fellow research groups who also wish to conduct remote assessment of cognition, or other factors such as sleep and physical activity, in HD or other cognitively compromised populations. See Figure 1 for a visual flow-chart of key steps taken during protocol development. This chapter also includes results from Study One of the thesis, wherein we addressed our first research aim, which was to demonstrate that people with the expanded HD repeat could complete the protocol (i.e., feasibility, acceptability). We used the findings from this chapter to update and improve our remote assessment protocol before beginning the studies presented in Chapters Three to Five.

### 2.2. Development of HD-Mobile and Remote Assessment Methodology

The final version of the remote assessment protocol utilised in the studies presented in Chapters Two to Five of the thesis included three cognitive tasks and four questionnaires which were administered in HD-Mobile and use of Fitbit One sleep and activity monitors. See Table 1 for a summary of these measures. As part of participant eligibility screening and

recruitment, participants also completed an online self-report demographic and symptom severity questionnaire. We considered several cognitive tasks, questionnaires, and actigraphy devices for inclusion in the protocol, but not all were included.

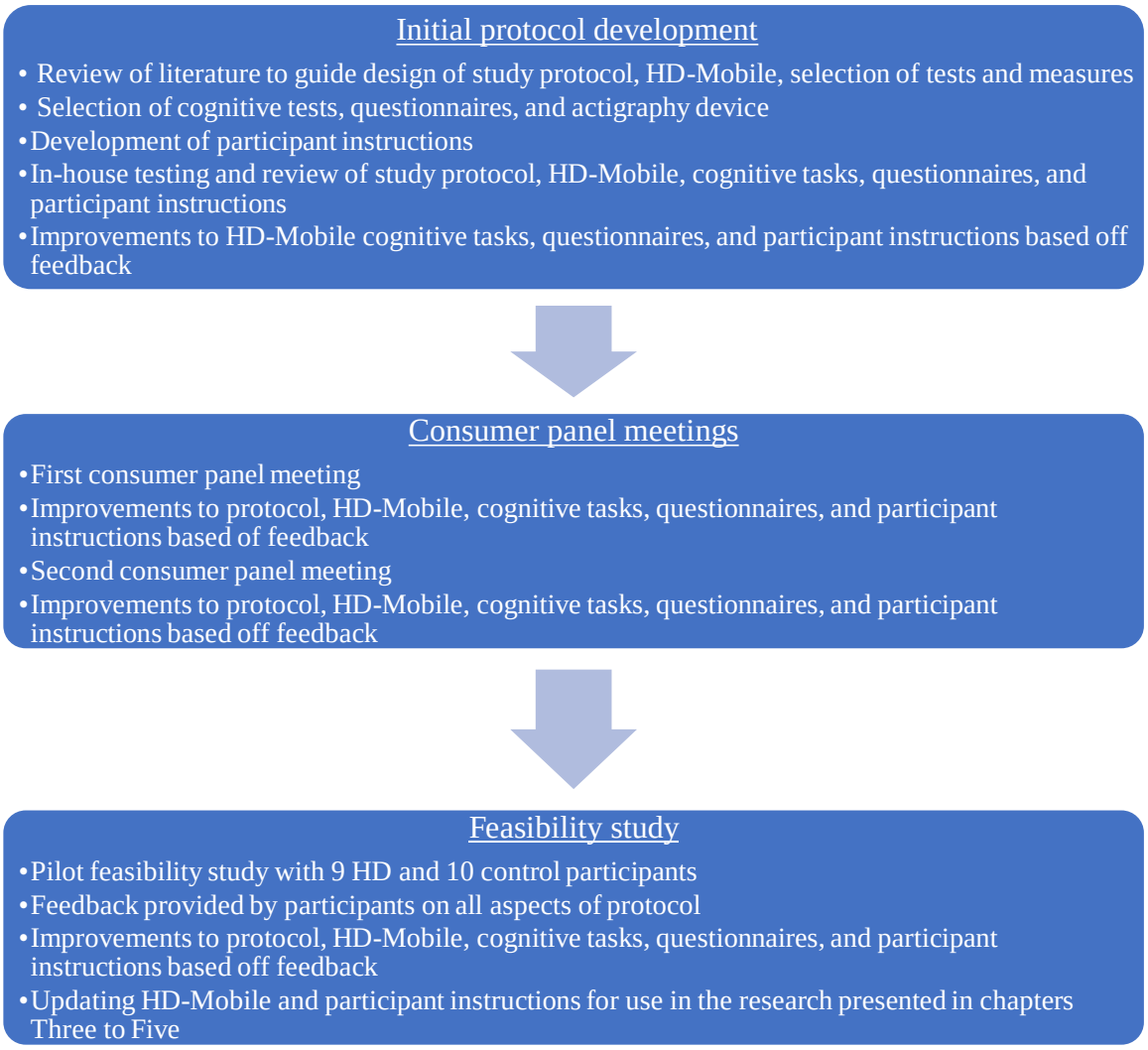



*Figure 1*. Flow chart summarising the key steps undertaken in developing HD-Mobile and all elements of the study protocol utilised in in the studies presented in Chapters Two to Five.

Participant burden was a primary consideration in selecting the final suite of cognitive tasks, questionnaires, and activity monitoring device. We aimed to collect as much scientifically robust, relevant, and useful information as needed to answer our research questions, with the minimal possible time commitment and burden to participants. We considered burden related to but separate from time commitment. For example, a

questionnaire with a small number of questions may have a minimal time commitment, but significant burden if the questions were complex and difficult for participants to understand and answer. A brief account of the selection of cognitive tasks, questionnaires, and selection of the Fitbit One over other sleep and activity monitoring devices is provided.

Table 1.

*Summary of Measures Used in the Thesis Research Project*

| Name of Measure | Method of Measurement | Domain assessed |
|---|---|---|
| Speeded Tapping | Computerised cognitive task | Cognition (psychomotor speed and coordination) |
| Two-Choice Discrimination | Computerised cognitive task | Cognition (decision making) |
| Object Information and Location Learning (OILL) | Computerised cognitive task | Cognition (visual recognition memory) |
| Daily Mood Sleep and Activity Questionnaire (DMSAQ) | Self-report questionnaire | Mood, sleep, physical activity |
| Hospital Anxiety and Depression Scale (HADS) | Self-report questionnaire | Mood |
| Scale of Outcomes in Parkinson's disease – Sleep scale (SCOPA-SLEEP) | Self-report questionnaire | Nighttime sleep and daytime sleepiness |
| Karolinska Sleepiness Scale (KSS) | Self-report questionnaire | Situational sleepiness |
| Fitbit One | Body worn accelerometer | Sleep and physical activity |

**2.2.1. Selection and development of cognitive tasks.** We selected the cognitive tasks for HD-Mobile with several criteria in mind. First, the tasks had to be relevant to cognitive domains known to be impaired in HD and to be suitable for completion on a smartphone. They also had to be easy to learn and understand with minimal instructions as participants would be completing tasks without the presence of an examiner. Ideally, the tasks would be relatively quick to complete, but sensitive to cognitive impairments in HD. By meeting these requirements, we could minimise participant burden and time commitment while collecting enough data to answer our research questions. We began the process of cognitive task selection with three candidate cognitive tasks. The tasks, Speeded Tapping, Two-choice discrimination, and a prospective memory task, were candidate
tasks for another cognitive assessment smartphone application, the Brain Health App, which our group was developing at the same time. The Brain Health app was for use in a healthy ageing study with the tasks selected to be sensitive to the onset of dementia. HD and several other forms of dementias such as Alzheimer's disease, have some overlapping areas of cognitive impairment as the diseases progress (Lezak, Howieson, Bigler, & Tranel, 2012). Hence, we considered the tasks included in the Brain Health App as good candidates for inclusion in HD-Mobile because they were designed for smartphones and would likely assess cognitive domains affected by HD.

Of the three Brain Health App tasks, we retained Speeded Tapping and Two-choice Discrimination. We retained Speeded Tapping for use with HD-Mobile as it is computerised, sensitive to symptom progression in HD, can be understood and completed by participants with manifest-HD, and clear deficits in performance are seen in manifest- and premanifest-HD samples when compared to healthy control samples (Stout et al., 2014; Tabrizi et al., 2012). We retained Two-choice Discrimination as although this style of task had not previously been used in HD research, related tasks have been successfully used with and

shown to be sensitive to cognitive changes in Parkinson's Disease, which is also a subcortical neurodegenerative movement disorder (Herz et al., 2018; Huang et al., 2015). Two-choice Discrimination style tasks generate large amounts of data which can be modelled mathematically resulting in detailed information about cognitive processes that determine performance on the tasks (Voss et al., 2013). Additionally, the variables obtained from modeling of Two-choice tasks can be mapped on to neural substrates (Frank, 2006; Frank et al., 2007; Herz et al., 2018; Huang et al., 2015), including the striatum which is a primary site of atrophy in HD (F. O. Walker, 2007). We chose not to retain the prospective memory task as the format of the task suited a longer-term longitudinal research approach than the two-day and eight-day study periods that we anticipated using for for the research project.

In place of the prospective memory task, we included Object Identification and Location Learning (OILL), a visual recognition memory task. The task was based on the paradigm developed by Buffalo, Bellgowan, and Martin (2006) and adapted from the version used by (Glikmann-Johnston, Halag-Milo, Hendler, & Gazit, 2020) which was designed for use on tablet devices. Visual memory deficits exist in early HD and visual tasks with similar methodology to OILL can be completed by manifest-HD samples and detect impairments in their visual memory functioning (Begeti, Schwab, Mason, & Barker, 2016; Glikmann-Johnston, Carmichael, Mercieca, & Stout, 2019). Hence, we expected people with the expanded HD repeat would be able to understand and complete the OILL task and that OILL would be sensitive to visual memory symptom progression in HD. Because we modified the task for presentation on smartphones, we conducted in-house testing and development was optimise performance of the task. This included optimising the difficulty level to avoid or minimise floor and ceiling effects by adjusting the number of stimuli to be learned and by testing color and black-and-white versions of the task stimuli. The in-house testing indicated an optimal level of difficulty would be achieved with eight coloured stimuli. We retained this

format for thesis study protocol. We also sought feedback on task instructions, for all three tasks, from in-house testers and the participants from the study described in this chapter so that these could be optimised for clarity and conciseness.

**2.2.2. Selection of actigraphy device.** We considered several consumer grade activity monitoring devices for use within the research project. We did not use polysomnography (PSG), the gold standard of sleep measurement (Montgomery-Downs et al., 2012), because PSG is typically laboratory based (Tonetti, Pasquini, Fabbri, Belluzzi, & Natale, 2008) and the protocol and research project were remote in nature. Research grade actigraphy devices were not viable due to their much higher cost compared to consumer grade actigraphy devices (de Zambotti, Claudatos, Inkelis, Colrain, & Baker, 2015; Ferguson et al., 2015). However, research has demonstrated comparable levels of accuracy for measuring sleep and physical activity levels between consumer and research grade actigraphy devices (de Zambotti et al., 2015; Ferguson et al., 2015; Maskevich et al., 2017; Montgomery-Downs et al., 2012).

We consulted the research literature to aid in the choice of an actigraphy device and selected the Fitbit One for use in the thesis research project. Regarding the literature, Ferguson et al. (2015) assessed seven consumer grade activity devices (Fitbit One, Fitbit Zip, Jawbone UP, Misfit Shine, Nike Fuleband, Striiv Smart Pedometer, and Withings Pulse) for measurement of physical activity and sleep. Of these devices, they rated the Fibit One, Fitbit Zip and Withings Pulse as the best performing devices. The Fitbit Zip, however, does not measure sleep, so it was not considered for use. The Fibit One and Jawbone Up were assessed elsewhere in the literature. Fitbit One and Jawbone Up were evaluated with healthy participants by Montgomery-Downs et al. (2012), and de Zambotti et al. (2015), respectively, and Maskevich et al. (2017) assessed both devices with a HD sample. Both devices were found to overestimate total sleep time, as was observed by Ferguson et al. (2015). However,

overestimates of sleep time also occur with research grade actigraphy (Montgomery-Downs et al., 2012). Nevertheless, the Fitbit and Jawbone devices were rated as cost effective and viable alternatives to much more expensive research grade actigraphy devices (de Zambotti et al., 2015; Maskevich et al., 2017; Montgomery-Downs et al., 2012). Other than the Ferguson et al. (2015) paper, we found no other peer reviewed evaluation of the Withings Pulse at the time of our evaluation process. As more information was available on the Fitbit One and Jawbone Up, we considered both for use in the project. We chose the  Fitbit One chosen due to it having a lower cost and because it outperformed the Jawbone Up in the direct a direct comparison by Ferguson et al. (2015).

**2.2.3. Selection and development of questionnaires.** HD-Mobile includes three established and validated questionnaires and one questionnaire which we developed specifically for the study. See Appendix 1 for copies of each questionnaire. The established questionnaires are the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), the Scale of Outcomes in Parkinson's Disease – Sleep scale (SCOPA-SLEEP; Marinus, Visser, van Hilten, Lammers, & Stiggelbout, 2003), and the Karolinska Sleepiness Scale (KSS; Kaida et al., 2006). We included the HADS to allow us to examine the effects that anxiety and depression can have on sleep quality (Riemann, Berger, & Voderholzer, 2001) and cognitive functioning (Austin, Mitchell, & Goodwin, 2001) and to allow direct comparisons of anxiety and depression symptoms between participant groups. Both subjective (SCOPA-SLEEP scale, KSS, DMSAQ) and objective (Fitbit One) sleep measures were included because both are equally important in measuring sleep disturbance (Choi-Kwon & Jeon, 2017; Landry, Best, & Liu-Ambrose, 2015; Lang et al., 2013).

We included the KSS to measure current alertness (Kaida et al., 2006) and to enable consideration of the effects of alertness levels on participant's performance on HD-Mobile cognitive tasks. The SCOPA-SLEEP questionnaire was selected to provide a measure of

participant nighttime sleep quality and daytime sleepiness in the month prior to participation in the study. We included the questionnaire we developed, the Daily Mood, Sleep, and Activity Questionnaire (DMSAQ), so we would have a subjective measure of sleep and wake timing and physical activity levels over the week, to compliment actigraphy data, as well as a subjective measure of daily mood to compliment the retrospective measure of mood provided by the HADS.

### 2.3. Consumer Panel Meetings and Further Methodology Development

As part of development of the overall remote assessment methodology, we held two rounds of meetings with members of the HD community, which we termed 'consumer panel meetings'. collaboration between researchers, clinicians, and patients is an essential component of good research practice and for the successful design and implementation of healthcare (Coulter & Ellins, 2006; Leidy & Vernon, 2008; Ocloo & Matthews, 2016; Payne et al., 2011). In particular, collaboration and consultation with patients and participants is critical when researchers plan to ask people with cognitive impairments to self-administer cognitive tests (Hagelkruys et al., 2016). The consumer panels consisted of a small group of people, nine for round one and six for round two, who either had premanifest- or manifest-HD or were family members or carers of someone with HD. We held the meetings so that the panel members, some of whom would be the very people to participate in the thesis research project, could give input into the design of the project's protocol. We also aimed to ensure, pre-emptively, that the research project and its expected outcomes would be meaningful and of interest to the panel members, and by extension the HD community. We translated their suggestions and input from the first round of meetings into the protocol and we presented these adjustments to the panel, for their approval and further comment, at the second round of consumer panel meetings. Feedback from the second round of meetings was implemented prior to recruiting participants for the feasibility study described in this chapter. We held the

first round of meetings in October and November of 2016 and the second round of meetings in December of 2016. Two meeting dates were offered for each round of meetings to accommodate the various schedules of invitees. A brief summary of the goals and outcomes of each meeting are presented below.

**2.3.1. Goals and outcomes of first round of consumer panel meetings.** Our goals for the first round of meetings were to seek the input of panel members on several points of interest that were directly related to the protocol design. To achieve these goals, we posed several key questions to panel members. I present the specific goals and summarized learnings from panel member responses in Table 2.

During the meeting, panel members were shown a prototype version of HD-Mobile and provided several suggestions for improvements, such as increasing the size of fonts, buttons, and images in the app, and to reduce the length of some of the task instructions. Based on feedback from the panel members, the International Physical Activity Questionnaire (IPAQ; Hagströmer, Oja, & Sjöström, 2006) was excluded from HD-Mobile due to the length and complexity of some of the questions that it contained. We responded to the panel's suggestions and implemented changes to HD-Mobile and the planned research protocol before meeting again with the panel two months later.

**2.3.2. Goals and outcomes of second round of consumer panel meetings.** Our goals for the second round of meetings were to present the changes made to HD-Mobile and the study instructions and to ask panel members if they had any further suggestion for improvements to any aspects of the protocol. I present the specific questions posed to panel members in the second round of meetings in Table 3.

Table 2.

*Summary of Key Questions Posed to Consumer Panel Members at First Round of Consumer Panel Meetings and Responses Received*

| Questions for consumer panel | Response from consumer panel |
|---|---|
| How confident would members of the HD community be in independently downloading and installing smartphone applications? | Most people with premanifest- and early manifest-HD should be able to confidently use smartphones independently. Those in later stages of the disease may need some assistance from close family members or the study investigators. |
| What level of time commitment and involvement do panel members feel would acceptable for them and other HD community members when participating in a research project? | Most people in their community would be accepting of spending five to fifteen minutes per day interacting with HD-Mobile over the course of the week. This was the planned time commitment we were considering requesting of participants in the studies presented in Chapters Three to Five. |
| Which format would HD community members most like to have study instructions delivered to them (e.g., written | The panel suggested that both written and video versions of the study instructions should be provided. |

| | |
|---|---|
| instructions, video-based instructions, or both)? | |

The panel members responded positively to the changes made to HD-Mobile from the first meeting. They specifically liked the reduced text in task instructions and stated the menus and content were well laid out. They also gave positive feedback to the inclusion of practice trials for some of the cognitive tasks. We provided participants with copies of the written study instructions and played sections of the video instructions to the group, to which they provided feedback. We implemented suggested changes and updated the instructions prior to launching the feasibility study.

Table 3.

*Summary of Key Questions Posed to Consumer Panel Members at Second Round of Consumer Panel Meetings and Responses Received*

| Questions for consumer panel | Response from consumer panel |
|---|---|
| How confident are panel members that they could engage with and understand HD-Mobile, Fitbit One, and the overall protocol based upon changes made? | Panel members felt confident that HD affected participants should be able to complete the study independently or with some support from family members or from the researchers conducting the study. |
| Do panel members suggest any further changes to HD-Mobile, study instructions, or any aspect of the protocol? | The language used was too technical in places and recommended reducing the use of acronyms in the video and written instructions. They felt both instructions types should begin with a clear overview of what was required on each day of the study |

| | |
|---|---|
| How confident are panel members that they or family their members could follow and understand the written and video instructions? | Panel members liked that the video and phone instruction included vision and photos of operation of HD-Mobile on a phone and operation of the Fitbit One. Notwithstanding some suggested changes, they believed people with the expanded HD repeat would be able understand and follow the instructions. |

### 2.4. Feasibility Study of HD-Mobile and Remote Assessment Protocol

**2.4.1. Objective and aims.** Our overall objectives for the feasibility study were to ensure that our proposed remote assessment protocol, which required independent operation of HD-Mobile and a Fitbit One device, could be completed by premanifest- and manifest-HD participants and to gain further direct input on how to improve the protocol. To achieve these objectives we recruited a small sample of participants (nine HD people with the expanded HD repeat and ten healthy control participants) and asked them to complete a truncated, two-day, version of our study protocol, in which they would perform all aspects of HD-Mobile and would operate the Fitbit One. We then conducted a semi-structured interview in which we asked participants to provide ratings on how confident they felt completing all aspects of the protocol (e.g., how confident they felt completing the questionnaires, completing the cognitive tasks, operating the Fitbit) and invited participants to suggest improvements we could make to any aspect of the protocol. Inviting feedback from participants on study protocols a key aspect of patient focused research design (Leidy & Vernon, 2008). We assessed completion rates and type and quantity of any data loss. We hypothesised there would be no difference between participants with expanded HD repeat and healthy control

participants in their confidence to independently complete HD-Mobile cognitive tasks, operate Fitbit One devices, or to complete all required aspects of their participation. We also hypothesised any data loss would be within acceptable limits.

### *2*.4.2. Method.

***2.4.2.1. Participants.*** Nine participants with the expanded HD repeat, five with premanifest-HD and four with manifest-HD ($M_{Age}$ = 47.2, SD = 5.54) and ten healthy control participants ($M_{age}$ = 44.64 SD = 8.5) were involved in the feasibility testing. All participants were required to be residents of Australia, fluent in English, 30 – 65 years of age, and to have access to an Android or Apple smartphone running Android 4.0 or iOS 7 (or above) above operating systems. Exclusion criteria were history of neurological illness (other than HD), psychiatric disorder, significant head injury, drug abuse, or current participation in any clinical drug trial. Participants with manifest-HD were further required to have a diagnosis of HD from a neurologist or neuropsychiatrist. Premanifest-HD participants were required to have a CAG expansion of at least 39 repeats and a disease burden score (DBS) of ≥ 200 (calculated as age x [CAG – 35.5]; Penney Jr, Vonsattel, Macdonald, Gusella, & Myers, 1997). Participants were recruited through the 'ENRU-STOUT' participant database at Monash University (managed by the Georgiou-Karistianis Experimental Neuropsychology Research Unit and the Stout Group), which is a database of people who have consented to being contacted about research participation opportunities.

***2.4.2.2. Materials.*** See Table 1 for the full list of measures and devices that participants used and completed during each of the studies presented in the thesis, including the study presented in this chapter. Each of the questionnaires, the cognitive tasks, and the Fitbit devices are described in detail in subsequent chapters. So, to avoid repetition, and because I do not analysed from any of the measures in this chapter, I have not presented further details on the thesis research materials here.

***2.4.2.3. Procedure.*** Following recruitment, participants were provided with all information and materials needed to complete study set-up and preparation. Participants were sent an email containing the HD-Mobile app (Android users), or instructions for downloading from the Apple Store (iPhone users), as well as username and password details for HD-Mobile and for their pre-created Fitbit account. Packs containing a Fitbit One and associated accessories, printed instructions, a USB flash drive containing video and written instructions, and another copy of username and password information were sent to participants' residential address. The written and video instructions informed participants of what was required on each day of the study as well as providing directions for downloading and installing HD-Mobile and the Fitbit app on their phones. Contact information of research staff was provided so participants could request assistance if needed. Participants could select a day to begin the study and were instructed to set up all apps and equipment on the day before beginning participation.

Active participation took place over a two day, approximately 48-hour, period. We instructed participants to begin wearing the Fitbit and to log in to HD-Mobile and complete the Day 1 tasks, within an hour of waking on Day 1. Before accessing the Day 1 tasks, participants were required to read the explanatory statement and complete the consent form from within HD-Mobile. Day 1 tasks included the SCOPA-SLEEP scale, the HADS, and the KSS. On the morning of Day 2, within an hour of waking, participants were required to log into HD-Mobile and complete the Day 2 tasks, which consisted of the DMSAQ, the KSS, and the cognitive tasks. Participants continued wearing the Fitbit for all of Day 2 until the morning of Day 3. On Day 3 participants could remove the Fitbit and return the device and accessories in a pre-paid return envelope. Following completion of active participation, participants were then contacted via phone, at a time of their choosing, and completed the structured interview (see Appendix 2 for the content of the structured interview).

***2.4.2.4. Data analyses.*** Statistical analysis was performed using IBM SPSS V.24 software. Missing data were excluded listwise from the analyses. Alpha was set at .05 for all analyses. We used two methods to determine how successfully participants could complete the study protocol independently. First, following completion of the study, including completion of all steps involving HD-Mobile and usage of the Fitbit for 48 hours, participants gave ratings, via structured interview, of their confidence when completing various aspects of the protocol. Ratings were given in a range of zero to ten, with zero being no confidence and ten being full confidence, and were calculated for the whole sample and were also separated and compared by group. We investigated confidence data for group level differences using independent samples t-tests. Second, we quantified study completion rates and types and kinds of any data. We did this as a less subjective means of understanding how successfully participants could complete the study independently. For example, a participant may have felt confident in their ability to operate the Fitbit but may have made operator errors leading to their sleep data not being recorded by the device, indicating their confidence, to a degree, may have been misplaced. This also acted as a check of the reliability of data transfer between HD-Mobile and our database and reliability of the Fitbit.

## 2.5. Results

Overall, participants generally felt confident in completing all aspects of the study. Mean ratings for all aspects of the study, for both the HD and control groups, were 7.67/10 or higher (see Figure 2), indicating moderate to high levels of confidence. The lowest mean rating, $M = 7.67$, $SD = 1.64$, range = 5.00 – 10.00, was given by the HD group for confidence in setting-up and using the Fitbit. The highest ratings given by the HD group were confidence in completing the questionnaires in HD-Mobile, $M = 8.50$, $SD = 0.80$, range = 7.00 – 9.50, and for the usefulness and informativeness of the written instructions, $M = 8.50$, $SD = 1.12$, range = 8.00- 10.00. A similar pattern of results was seen for the control group whose lowest

rating was also related to the Fitbit, this time the rating of the overall usability of the Fitbit, *M* = 8.28, *SD* = 2.08, range = 4.00 – 10.00 and, as with the HD group, the highest rating was for confidence in completing the questionnaires in HD-Mobile, *M* = 8.90, *SD* = 1.22, range = 6.00 – 10.00. Apparent in Figure 7 is a slight trend for control participants to rate the confidence in completing individual aspects of the study slightly higher than HD participants. However, with this small initial sample, there were no significant differences (all *ps* ≥ .46) and both HD, *M* = 8.06, *SD* = 1.51, range = 6.00 – 10.00 and healthy control groups *M* = 8.45, *SD* = 1.17, range = 6.50 – 10.00, were similarly confident in their ability to complete the study independently following working through the study instructions.

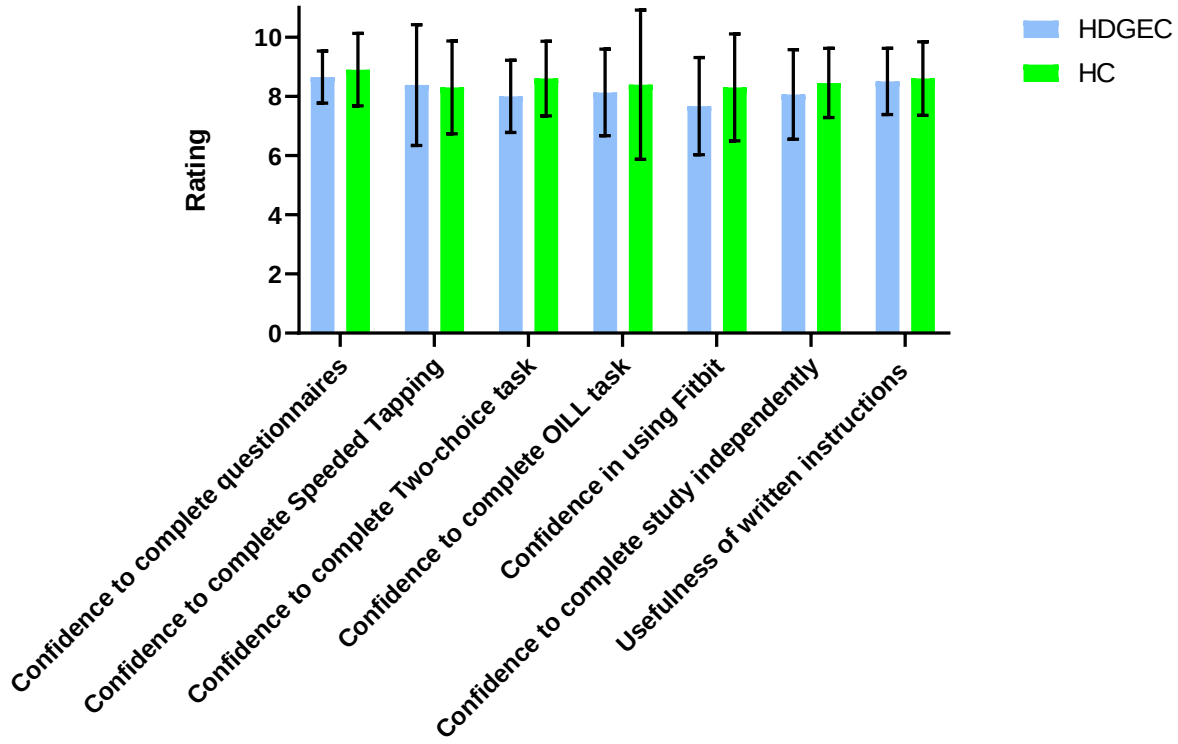



*Figure 2.* Mean and 95% CI for participant ratings of confidence to independently complete various aspects of the study and on usefulness of the written instructions. Note: Ratings were on a scale of 1 – 10 with 10 indicating highest confidence.

In addition to providing ratings on their confidence to complete the study, we asked participants to suggest improvements to the study protocol, such as suggesting how to

improve HD-Mobile or how to improve the participant instructions. We also sought comments for each individual rating so that reasons for poor and good rating could be understood and, together with other solicited suggestions, improvements made to the study protocol. Feedback from participants was mostly positive, with occasional suggestions for improvement. Only two participants provided a non-positive response, with these being in relation to how they felt about completing the study remotely on their smartphones: "personally I feel ambivalent" (HD participant) and "5/10 because HD is making it harder for me to do things. IQ seems intact, but processing is muddled". Table 4 provides a summary of participant responses for key aspects of the protocol. Of note, we found most participants used the written instructions and did not view the video instructions. Only five participants reported viewing the video instructions.

Table 4. Summary of participant feedback and responses for key aspects of the protocol.

| Aspect of protocol commented on | Comments and feedback provided |
|---|---|
| Written Instructions | "They were very informative, but there was a lot of information, so a summary or flow chart would be useful" (healthy control participant).<br>"I was confused when to start and looked at the video for help. It wasn't clear. A timeline or breakdown would help" (healthy control participant). "Straightforward and easy to use" (HD participant). |
| Video instructions | "Easy to use. I referred to the video instructions when I couldn't understand the written instructions" (healthy control |

| | |
|---|---|
| | participant). “Really good” (HD participant). |
| Object Identity and Location Learning task | “The example had three images and the actual one had eight. I wasn’t ready to memorise eight to ten pictures” (healthy control participant).<br>“You have to really read it [the prompts] to pick up if you needed to remember object or location. People might miss that” (healthy control participant).<br>“It was a little confusing. I was focusing on the pictures rather than what was being asked [whether to remember object identity or location]” (HD participant). |
| Two-choice Discrimination task | Several participants commented that they initially struggled on Two-choice Discrimination as their idea of “few” was five or six and was much less than what was expected from the task. |
| Completing the study remotely | “Really good. It is a good way to be able to do it. It takes 5 - 10 minutes to do” (HD participant).<br>“Convenience is awesome”, and “No problem. It was easy and the reminders were helpful” (healthy control participant). |

| | |
|---|---|
| | "It was very natural; I use my phone a lot" (HD participant).<br>"Personally, I feel ambivalent as I have now done so many studies" (HD participant)<br>"5/10 because HD is making it harder for me to do things. IQ seems intact, but processing is muddled" (HD participant). |

Regarding study completion rates, all participants who enrolled in the study completed Day 1 and Day 2 tasks and made themselves available for the structured interview at the end (see Figure 3 for summary of completion rates and data loss). All questionnaires and all cognitive tasks were completed in HD-Mobile and all data was successfully transferred to our database for analysis. However, modest data loss occurred from the Fitbit: 81.58% of sleep data and 89.47% of activity data was obtained. One manifest-HD participant was unable to pair the Fitbit to their phone meaning no sleep or activity data was recorded for that participant. Another manifest-HD participant appeared to connect the Fitbit to the phone, but had no activity or sleep data recorded, indicating a fault with the device. Three healthy control participants forgot to set sleep mode on one of the two nights. However, one night of sleep data was recoverable by using the sleep diary completed in HD-Mobile to create a sleep log in Fitbit. One healthy control participant removed the Fitbit at night, not realising it was supposed to be worn to bed.

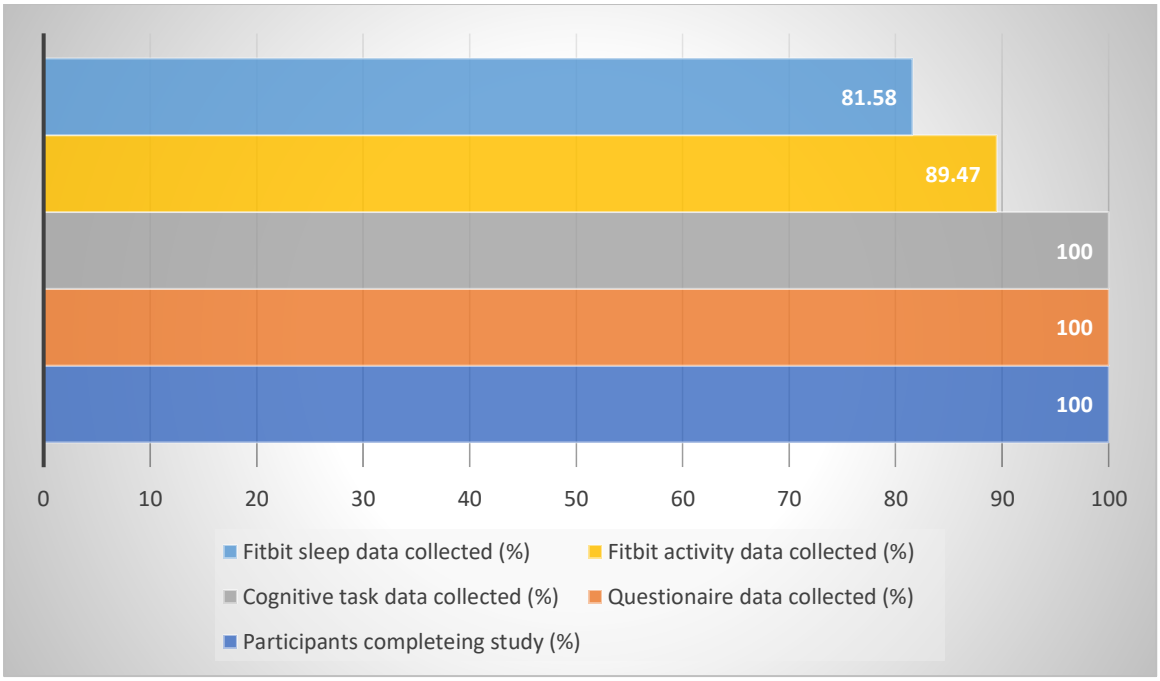



*Figure 3.* Percentages of participants completing the study after beginning and percentages of data types collected without loss. Note: $N$ =19.

### 2.6. Discussion

This is the first work we know of to attempt remote cognitive assessment in HD and also the first to simultaneously remotely assess cognition, sleep, and activity. Results demonstrate our remote assessment protocol, utilising HD-Mobile and Fitbit One activity monitors, is feasible for use within a HD sample, thus achieving our primary objective of the pilot feasibility study. That we were successful in this was evidenced by the consistently high mean ratings for confidence of participants to independently complete all aspects of the protocol, and most importantly that ratings did not differ between the HD and healthy control groups. If ratings provided by the HD group were lower than controls, it would indicate that some elements of the protocol were causing them specific additional burden or confusion, which would be problematic for plans to deploy the protocol in a large sample or project. We would also have failed to meet key recommendations from the literature that computerised cognitive assessment batteries be designed with any cognitive or technical literacy deficits of

target samples in mind (Bauer et al., 2012; Hagelkruys et al., 2016; Parsey & Schmitter-Edgecombe, 2013).

An additional key indicator of the feasibility and strength of design of the protocol was the 100% completion rate, demonstrating objectively that participants could successfully use HD-Mobile and the Fitbit activity device and found the process tolerable. Tolerability was also highlighted by the positive comments received from participants regarding how they felt about completing cognitive tasks and questionnaires on their smartphones, with only two of 19 participants responding in a non-positive manner. The feasibility of using of HD-Mobile was further supported by the finding that no loss of data occurred. All questionnaires and tasks were completed, and all data was successfully transmitted to our database. It is likely that these positive results were influenced by our decision to consult with HD-community members during development of the remote assessment protocol. This approach aligns with recommendations from the United States Food and Drug Administration (US Food and Drug Administration, 2017) and several authors (Coulter & Ellins, 2006; Leidy & Vernon, 2008; Ocloo & Matthews, 2016; Payne et al., 2011) that have advocated the value of patient perspectives on their conditions and available therapies and research endeavors related to their conditions. Congruently, the consumer panel identified several areas for improvement that our research group had not considered and would have likely remained in the protocol if not for consumer panel input. After the feasibility study, following participant suggestions, we were able to make yet further improvements to the protocol prior to the studies presented in subsequent chapters.

The loss of some sleep and physical activity data through the Fitbit caused some concern. The data loss was primarily due to user error and we identified two primary means of better supporting participants in Phase Two of the project. First, we adjusted the written and video instructions to provide clearer instructions on Fitbit operation and we also began

monitoring Fitbit usage more closely. Eventually, we learned that daily monitoring of Fitbit usage was needed as participants sometimes connected the Fitbit correctly, but forgot to log sleep on some nights or depleted the battery without realising, meaning data loss would occur if Fitbit use was only checked at the commencement of the study.

Our next major aim was to evaluate the quality of data generated from the HD-Mobile cognitive tasks, and to consider reliability and key forms of validity, such as known groups validity. This was an essential step given our overarching goal of employing our remote assessment methodology in large scales studies, such as those monitoring cognitive outcomes of clinical trials, as concern has been previously raised in the literature regarding a tendency for some computerised cognitive assessment measures to be insufficiently validated (Bauer et al., 2012; Schlegel & Gilliland, 2007; Wild et al., 2008). Thus, although we could be confident that large scale deployment of HD-Mobile was feasible, specific validation of the HD-Mobile cognitive tasks was needed. I present a validation study of the HD-Mobile cognitive tasks in the following chapter.

## Chapter Three: Study Two

## Feasibility and Initial Validation of 'HD-Mobile', a Smartphone Application for Remote Self-administration of Performance-based Cognitive Measures in Huntington's Disease

### 3.1. Explanatory Notes

In this chapter, I present Study Two from the thesis research. This chapter addresses our second research aim, which was to demonstrate that the cognitive tasks we developed for the research project would be reliable, and valid, across several forms of validity, and sensitive to cognitive impairment in HD compared to controls. To address this aim we compared the performance of premanifest-HD, manifest-HD, and healthy control participants on the HD-Mobile cognitive tasks. We also assessed how HD-Mobile cognitive outcomes, for participants with the expanded HD repeat, compared to pen-and-paper cognitive task data and to measures of symptom severity. The results indicated that the HD-Mobile cognitive tasks had robust test-retest reliability, known-group and concurrent validity, achieving our aim. The results of Study Two extended the results of Study One, by demonstrating that remote assessment of cognition in HD is not only feasible but can generate reliable, valid, and scientifically meaningful data.

**Note.** This chapter was published in the *Journal of Neurology* and is presented in its published form. Within the chapter is a reference to supplementary material. The supplementary material that was submitted with the study manuscript included content from Study One, presented in Chapter Two, and so is not reproduced later in this chapter or the appendix to the thesis.

**McLaren, B**., Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E.-C., Murray, N. W. G., Loy, C., Bellgrove, M. A., Stout, J. C. (2020). Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of

performance-based cognitive measures in Huntington's disease. *Journal of Neurology*.

Journal of Neurology
https://doi.org/10.1007/s00415-020-10169-y

ORIGINAL COMMUNICATION

# Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease

**Brendan McLaren[1] · Sophie C. Andrews[1,2,3] · Yifat Glikmann-Johnston[1] · Emily-Clare Mercieca[1] · Nicholas W. G. Murray[5] · Clement Loy[4,5] · Mark A. Bellgrove[1] · Julie C. Stout[1]**

Received: 15 May 2020 / Revised: 29 July 2020 / Accepted: 13 August 2020


## Abstract

**Objective** Smartphone-based cognitive assessment measures allow efficient, rapid, and convenient collection of cognitive datasets. Establishment of feasibility and validity is essential for the widespread use of this approach. We describe a novel smartphone application (HD-Mobile) that includes three performance-based cognitive tasks with four key outcome measures, for use with Huntington's disease (HD) samples. We describe known groups and concurrent validity, test–retest reliability, sensitivity, and feasibility properties of the tasks.

**Methods** Forty-two HD CAG-expanded participants (20 manifest, 22 premanifest) and 28 healthy controls completed HD-Mobile cognitive tasks three times across an 8-day period, on days 1, 4, and 8. A subsample of participants had pen-and-paper cognitive task data available from their most recent assessment from their participation in a separate observational longitudinal study, Enroll-HD.

**Results** Manifest-HD participants performed worse than healthy controls for three of four HD-Mobile cognitive measures, and worse than premanifest-HD participants for two of four measures. We found robust test–retest reliability for manifest-HD participants (ICC = 0.71–0.96) and with some exceptions, in premanifest-HD (ICC = 0.52–0.96) and healthy controls (0.54–0.96). Correlations between HD-Mobile and selected Enroll-HD cognitive tasks were mostly medium to strong ($r$ = 0.36–0.68) as were correlations between HD-Mobile cognitive tasks and measures of expected disease progression and motor symptoms for the HD CAG-expanded participants ($r$ = −0.34 to −0.54).

**Conclusions** Results indicated robust known-groups, test–retest, concurrent validity, and sensitivity of HD-Mobile cognitive tasks. The study demonstrates the feasibility and utility of HD-Mobile for conducting convenient, frequent, and potentially ongoing assessment of HD samples without the need for in-person assessment.

**Keywords** Huntington's disease · Cognition · Remote assessment · Neuropsychology

## Introduction

Cognitive impairments in Huntington's disease (HD) often appear many years before clinical diagnosis [1, 2] which, according to traditional diagnostic criteria for HD, is based on the emergence of significant observable motor

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00415-020-10169-y) contains supplementary material, which is available to authorized users.

✉ Julie C. Stout
Julie.stout@monash.edu

[1] Turner Institute for Brain and Mental Health. School of Psychological Sciences, Monash University, 18 Innovation Walk, Clayton, VIC 3800, Australia

[2] Neuroscience Research Australia, Sydney, NSW, Australia

[3] School of Psychology, University of New South Wales, Sydney, NSW, Australia

[4] Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia

[5] Huntington Disease Service, Westmead Hospital, Sydney, NSW, Australia

Published online: 03 September 2020

dysfunction [3, 4]. The impact of cognitive symptoms on functioning and quality of life in HD extends beyond the impact of motor symptoms alone [5, 6]. Because exciting new potential treatments for HD are currently in clinical trials [7], the current need is even greater for efficient, cost-effective methods of phenotyping people with the HD gene in large-scale studies. Such studies are important to enable future investigations of additional genetic modifiers and for monitoring response to therapeutics. For example, recent studies investigating genetic modifiers of age of onset of cognitive symptoms required sample sizes in the thousands, but because data in these studies were collected using face-to-face assessments with clinicians, the acquisition of sufficient datasets took many years, limiting the rate of progress [8, 9].

Remote, self-administered, performance-based cognitive tasks may provide a solution to the problem of labor-intensive face-to-face assessments. For example, a mobile app-based approach could enable rapid, inexpensive, accurate, and convenient means of collecting cognitive data [10]. Furthermore, remote computerized (e.g., smartphone-based) methods increase the inclusiveness of participation by improving accessibility for people who are geographically dispersed [11–13]. Additionally, self-administered tasks can enable frequent repeat testing, which is ideal for monitoring cognitive change over time [13, 14], and by extension, ideal for monitoring day-to-day fluctuations in functioning that would be invisible in, for example, an annual clinic visit [15–17]. Thus, several compelling reasons support the use of remote self-assessment of cognitive abilities. Remote assessment may complement traditional clinic-based assessment, or, when clinic-based assessments are impracticable, may be an alternative that can provide an indication of function where no other assessments would be possible. Such a requirement may be best served by cognitive assessments on smartphones, which are now a ubiquitous technology that could be utilized at minimal cost under a 'bring your own' model.

We have created a 'bring your own' mobile app, HD-Mobile, designed for use on smartphones and for assessment of functioning in HD. In this first step, we tested HD-Mobile using three computerized performance-based cognitive assessment tasks, including Speeded Tapping, Two-choice Discrimination, and Object Information and Location Learning (OILL), which allow assessment of several cognitive domains in a brief testing session. Our aim was not to conduct a full cognitive battery remotely. Instead, at this stage of research, we wanted to explore cognitive assessments that would be able to be performed without intervention of a clinician, and that would include tests across a range of response requirements, including selecting correct answers, choosing between alternatives, and performing a speeded motor task. A brief assessment is particularly important for success in the context of participants working independently without clinician support and needing to incorporate the assessment into their daily life [10]. Speeded Tapping is a measure of psychomotor speed and motor control that is sensitive to disease stage and symptom progression in HD [2, 18]. Two-choice discrimination tasks are computerized tasks that assess binary-choice decision making, can be modeled mathematically [19], and are sensitive to cognitive impairment in Parkinson's disease, another sub-cortical neurodegenerative disorders [20, 21]. Object Information and Location Learning [22] is a task with two visual learning and recognition conditions that examine hippocampal-dependent memory. Research with other visuospatial memory-based tasks has identified deficits in visuospatial-based memory performance in manifest-HD participants [23, 24]. Taken together, Speeded Tapping and visual memory-based tasks can be completed by manifest-HD participants and are sensitive to cognitive impairment in HD, and Two-choice discrimination tasks have been successfully used with other neurodegenerative samples. Considering this, we included these three cognitive tasks in HD-Mobile as we expected the tasks to assess several domains of cognitive functioning which are affected in HD while being brief to complete and simple enough for participants to self-administer in the absence of a clinician.

Remote self-assessment using cognitive measures will be of limited value until their feasibility and validity are established [11, 16, 25]. Hence, the aim of this study was to determine the feasibility of remote self-administration, without any face-to-face contact or in-person training, using cognitive tasks in people with the HD gene expansion (i.e., HD gene expansion carriers; HDGEC) and healthy control participants. We also aimed to assess the known-groups and concurrent validity, sensitivity to indicators of disease progression, and test–retest reliability of the three HD-Mobile performance-based outcome measures.

## Method

### Participants

We studied 69 participants, including 42 genetically confirmed HD gene-expansion carriers and 28 healthy controls (Table 1). Twenty-two HD gene-expansion carriers were in the premanifest stage of HD (premanifest-HD) and 20 were diagnosed and in the early stages of manifest HD (manifest-HD). Inclusion criteria were being between 18 and 70 years of age, fluency in English and having access to an Android or Apple smartphone running Android 4.0 or iOS 7 (or above) operating systems. Exclusion criteria included current participation in any clinical drug trial, history of neurological illness (other than HD), psychiatric

**Table 1** Participant characteristics

| Characteristic | Healthy controls | Premanifest-HD | Manifest-HD |
|---|---|---|---|
| *n* | 28 | 22 | 20 |
| Age [mean years (SD), range] | 44.43 (11.11)<br>28–65 | 42.09 (8.52)<br>21–58 | 51.60 (8.03)<br>31–65 |
| Education [mean years (SD), range] | 14.80 (1.66)<br>12–19 | 13.25 (2.09)<br>10–18 | 14.33 (2.14)<br>12–18 |
| CAG repeat length [mean (SD), range] | – | 42.95 (2.16)<br>41–51 | 42.88 (2.11)<br>41–49 |
| Disease-burden score [mean (SD), range] | – | 304.12 (56.13)<br>217.50–435.00 | 379.64 (58.85)<br>275.00–484.50 |
| CAG-age product (CAP) [mean (SD), range] | – | 382.47 (66.45)<br>270.86–541.72 | 475.62 (60.78)<br>367.00–589.38 |
| HADS anxiety [mean (SD), range] | 6.07 (4.60)<br>0–16 | 6.36 (4.29)<br>0–15 | 5.85 (3.62)<br>2–17 |
| HADS depression [mean (SD), range] | 3.14 (3.48)<br>0–16 | 3.50 (3.26)<br>0–11 | 5.40 (4.45)<br>0–14 |
| UHDRS TFC (self-report) [mean (SD), range] | 12.96 (0.19)<br>12–13 | 12.64 (0.90)<br>9–13 | 9.26 (2.42)<br>6–13 |
| *n* | – | 12 | 7 |
| UHDRS TMS [mean (SD), range] | | 1.25 (2.55)<br>0–7 | 15.89 (12.59)<br>1–40 |

*HADS* Hospital Anxiety and Depression Scale, *UHDRS TFC* Unified Huntington's Disease Rating Scale Total Functional Capacity, *UHDRS TMS* Unified Huntington's Disease Rating Scale Total Motor Score, which was only available for participants involved with Enroll-HD

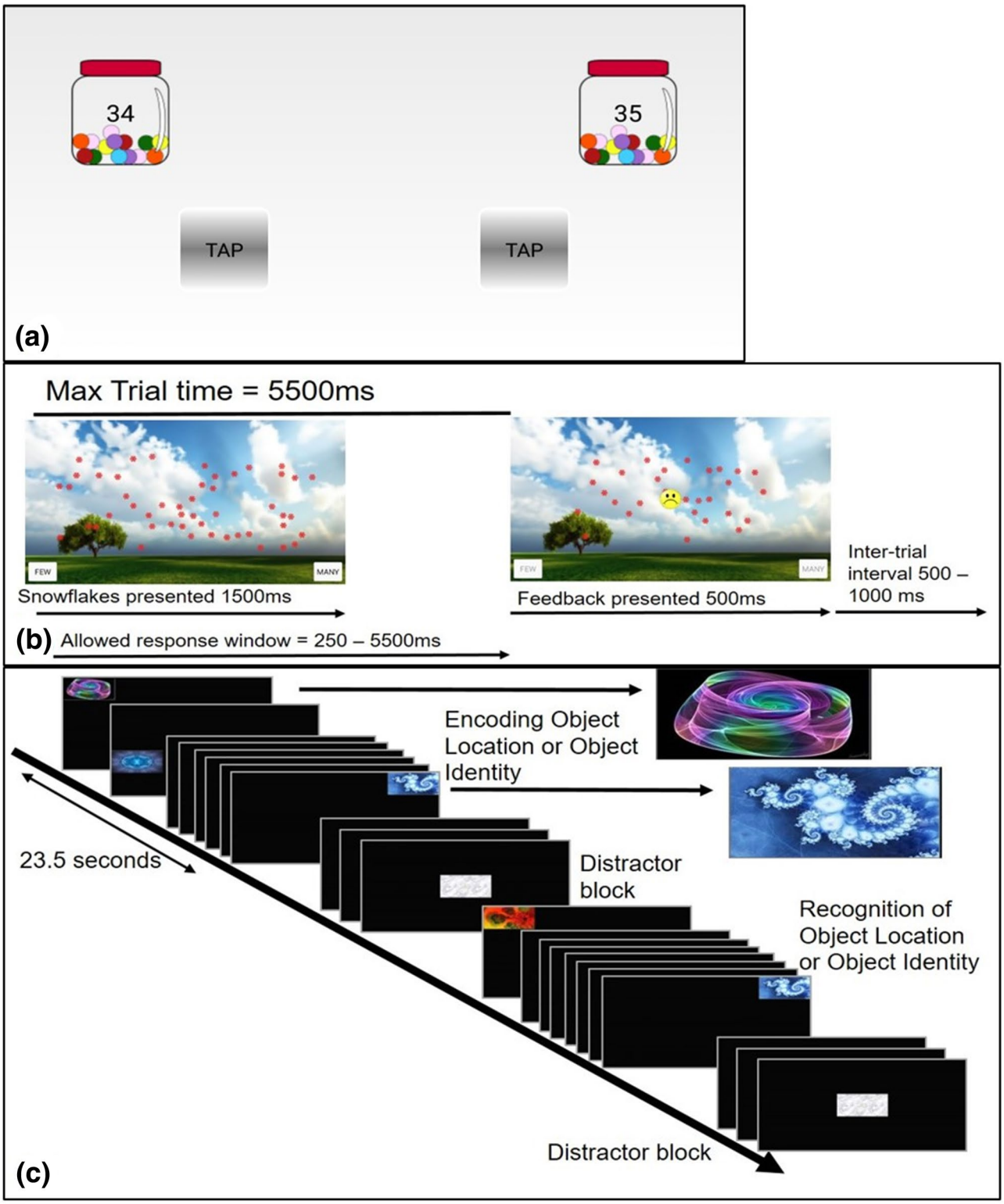
disorder, significant head injury, or drug abuse. Manifest-HD participants had all been diagnosed by a neurologist or neuropsychiatrist. For the premanifest-HD group, CAG expansions were 41 or more, and disease-burden score (DBS) was ≥ 200 (calculated as age × [CAG − 35.5]) [26]. We also calculated the CAG-Age product for HD gene-expansion carriers (calculated as age × [CAG − 33.6]) [27]. As is typical in HD samples, the manifest-HD group was significantly older than the premanifest-HD and healthy control groups ($p = 0.005$ and $p = 0.033$, respectively). The control group had a higher level of education than the premanifest-HD group ($p = 0.018$). The manifest-HD group had significantly higher Unified Huntington's Disease Rating Scale (UHDRS) total motor score (TMS), DBS, and CAG-Age product scores than the premanifest-HD group (all $ps < 0.008$), and lower UHDRS Total Functioning Capacity (TFC) score than the premanifest-HD and healthy control groups (both $ps < 0.001$). There were no other significant differences between groups (all $ps > 0.06$). One healthy control participant self-reported significant levels of anxiety and depression and one manifest-HD participant reported significant anxiety. These participants were not excluded from analyses. For a subset of analyses in which we examined convergence of performance on HD-Mobile with a separate ongoing study (Enroll-HD), we included all nineteen participants, (manifest-HD = 7, premanifest-HD = 12) for which such data were available, and chose their most recent Enroll-HD assessment for these analyses.

We recruited participants through participant databases at Monash University, Melbourne, and Westmead Hospital, Sydney. Additional participants were recruited using advertisements for the study posted on Facebook and at local HD-community events.

## Materials

HD-Mobile is a smartphone application developed by our group for touch screen smartphones operating Android 4.0 and iOS 7 and above. HD-Mobile can send reminder messages to participants, administer cognitive tests and self-report questionnaires, and automatically transmit data to the lab server.

For this study, HD-Mobile included three performance-based cognitive tasks; Speeded Tapping, Two-choice Discrimination, and Object Identity and Location Learning (OILL). Briefly, Speeded Tapping is a bimanual speeded finger-tapping task. For this task, participants viewed the mobile phone display which showed two lolly jars (see Fig. 1a). Using onscreen instructions, participants were asked to 'fill the jars' by tapping the response 'TAP' buttons on the screen as fast as possible with alternating thumbs. The task comprised three blocks, each completed with 50 correct (alternated) taps. The main dependent variable was mean inter-tap interval (milliseconds), termed "tapping speed". Following inverse log transformation to correct for non-normality, higher values represented faster tapping speed.




The Two-choice Discrimination task (Fig. 1b) was modelled after the Numerosity Discrimination task [19]. Participants chose rapidly between two alternatives, with limited information to guide their choices, and used feedback to guide future choices. Participants were presented with a scene depicting snowflakes and were instructed to quickly decide, without counting, whether there were "few" or "many" snowflakes on the screen. After each decision, feedback was provided in the form of a happy (correct choice) or sad face (incorrect choice). Three blocks of forty trials

◂**Fig. 1** Screenshots and schematic diagrams of HD-Mobile cognitive tasks. An example screenshot of Speeded Tapping (**a**) mid-way through the task is shown. Each Two-choice Discrimination (**b**) trial lasted for between 250 and 5500 milliseconds (ms). Snowflakes were presented for up to 1500 ms, at which point they disappeared from the screen to prevent counting. Participants then had a further 4000 ms to respond. A response within 250–5500 ms triggered feedback (a happy or sad emoticon) for 500 ms followed by a 500 ms inter-trial interval. If participants did not respond within the response window a 1000 ms inter-trial interval was used. Response buttons were disabled during feedback and the inter-trial interval. The OILL task (c) involved presentation of colored fractals on a black background in one of 16 possible locations on an invisible 4×4 grid. The encoding phase was 23.5 s in duration, with each of eight fractals presented for 3000 ms, followed by 500 ms gap between each fractal. During distractor and recognition trials, stimuli were presented on a screen until a response was provided by participants

were completed. The dependent variable was overall accuracy (percent correct), with higher scores indicating better performance.

The Object Identification and Location Learning (OILL) task was based on the paradigm developed by Buffalo and colleagues [28] and adapted from the version used by Glikmann-Johnston et al. [22]. The task assessed visual memory in two conditions, Object Identity and Object Location. Each condition occurred twice and included three block types: learning, distractor, and recognition (see Fig. 1c). The learning trials involved sequential presentation of eight unique color fractal objects in separate locations on the screen with participants instructed to remember either the objects themselves (object identity condition) or the locations of the objects (object location condition). The distractor block was a letter identification task. The recognition block tested participants' recognition of the object (object identity condition), or the location (object location condition) by presenting them with a series of fractals to which they had to select 'yes' or 'no' buttons displayed on the screen. Overall accuracy (percent correct) for object identity and object location trials was recorded and used in subsequent analyses. Higher scores indicated better performance.

In addition, HD-Mobile included a series of questionnaires assessing mood, sleep, and activity and participants wore Fitbit One activity/sleep monitors. Of these, only a mood measure, Hospital Anxiety and Depression Scale (HADS) [29], was inlcuded in the current report. Data from Fitbit devices and other questionnaires were used for a separate set of analyses with separate aims and are not reported here. The HADS is a widely used and reliable [30, 31] 14-item self-report questionnaire that assesses anxiety and depression. During recruitment, participants completed a self-report version of the UHDRS TFC [32], a five-item scale, which assesses functional capacity and has acceptable accuracy when self-completed [33]. Participants provided ratings of their own ability to complete the daily activities described by the scale. The questions, along with all demographic questions, were completed via an online-questionnaire before participants downloaded HD-Mobile. Scores range from zero to 13, with higher scores indicating greater functional capacity. The Enroll-HD cognitive battery includes several well-establised pen-and-paper cognitive tasks including Symbol Digits Modalities Test, F–A–S and Animals fluency subtests from the Controlled Oral Word Association Test, and the word-reading condition from the Stroop Color Word Test. We also had total motor score (TMS), a clinician derived measure of motor symptoms [32], data available for seventeen of the nineteen Enroll-HD participants.

## Procedure

Following recruitment, we provided participants with video and written instructions for the installation of HD-Mobile and described what they were required to do each day of the study. No face-to-face training was provided to participants as the aim was to assess if participants could complete the protocol entirely remotely with no provision of in-person training. Participation occurred over an 8-day period, during which participants completed at least one study procedure on each day, including completing the cognitive tasks on three separate occasions. We were limited to an 8-day study period due to the battery life of the Fitbit One devices. Selecting an 8-day study period allowed the devices to be mailed to participants and then worn for as long as possible with a minimal chance that participants would need to recharge the devices part-way through the study. We then spaced the cognitive assessment days out within the 8-day study period. At 9 am each morning of the study, participants would receive a notification on their phone from HD-Mobile reminding them to complete their daily tasks. On the morning of day 1, participants logged into HD-Mobile and completed an electronic informed consent form, the three cognitive tasks, all the questionnaires, and began wearing the Fitbit One. The cognitive tasks were completed again on days 4 and 8. Participants were asked to complete tasks within an hour of waking on days 1, 4, and 8. Participants continued wearing the Fitbit until the morning of day 8. The HD-Mobile app enforced the completion of a standard set of procedures in an order specified by us; for example, day 1 tasks could not be accessed until informed consent was completed and day 2 tasks could not be accessed until day 1 tasks were completed. Total time commitment required from participants was two and a half hours. Remarkably, we experienced no attrition in the study.

## Data analysis

Statistical analyses were performed using IBM SPSS V.24 software. Missing data were excluded pairwise from the

analyses. Alpha was set at 0.05 for all analyses. Mean of response times from Speeded Tapping was positively skewed and was, therefore, transformed using the reciprocal of the natural log. Our primary means of determining feasibility were via assessing study completion rates. We required an overall study completion rate of 90% by HD gene-positive participants to match adherence rates of a smartphone-based study in Parkinson's disease [34] and to be confident that most premanifest- and early manifest-HD patients could operate HD-Mobile and participate in remote assessment. Prior to this study, we developed the project to ensure feasibility in two ways. First, we met an advisory panel of nine members of the HD community who either had premanifest- or manifest-HD or were family members or carers of someone with HD. The advisory panel gave direct input into the remote assessment protocol, including giving direct feedback on the design of HD-Mobile and supporting materials, to ensure the project would be feasible. Second, we also conducted a pilot study ($n=5$ premanifest-HD, $n=4$ manifest-HD, and $n=10$ healthy control participants) to ensure feasibility. The HD gene-positive and healthy control participants were confident in using HD-Mobile, with no difference in confidence levels (see supplement for data from this study). Hence, for this study, we assessed completion rates to determine if the indications of feasibility obtained from the pilot study held for a larger sample.

We viewed the examination of known-group validity of HD-Mobile cognitive tasks from the point of view that the literature has clearly established differences in the magnitude of cognitive impairment between people with manifest-HD, premanifest-HD, and healthy control participants. People with manifest-HD have clear cognitive deficits compared to controls and people with premanifest-HD, who are nearing clinical onset, have milder cognitive deficits that may not be discernable versus healthy controls [2, 35]. Because all participants were recruited into the study with established group designations, based on their HD-status, to establish known-group validity, we required HD-Mobile tasks to be able to detect the significant cognitive impairments that are reliably present in manifest-HD groups when compared to healthy control groups. We did not compare HD-Mobile cognitive task performance to in-person collected baseline cognitive data because the group-level cognitive profiles of manifest- and premanifest-HD groups are reliably established and as we were deliberately not meeting with our participants. Known-group validity was examined using mixed between- and within-subject analyses of variance (ANOVA) with Tukey's Honestly Significant Difference post hoc test to analyze the effects of group (premanifest-HD, manifest-HD, healthy control) and day (1, 4, 8) on cognitive performance for each task.

To examine variability in performance across the days (test–retest reliability), we used the intraclass correlation coefficient (ICC) and set an ICC value of 0.70 as a benchmark for acceptable reliability [36]. To examine relationships between day 8 HD-Mobile cognitive task performance and results from Enroll-HD assessment (concurrent validity), as well as associations with TMS and DBS (sensitivity to indicators of disease progression), we used Pearson's correlation coefficients. We analyzed day 8 cognitive data for these analyses to limit the impact of practice effects. We utilized available Enroll-HD cognitive data in concurrent validity assessment. Although we were not attempting to approximate the Enroll-HD cognitive tasks, we expected participants performing poorly on Enroll-HD tasks would also perform poorly on HD-Mobile cognitive tasks, indicating concurrent validity with established measures of cognition in HD. We particularly expected agreement between tasks loading on similar domains (e.g., SDMT has psychomotor and visual memory components, Speeded Tapping has a psychomotor component, and the OILL task has a visual memory component). We set a medium effect size ($r \geq 0.30$) as an indicator of acceptable levels of concurrent validity and sensitivity. We assessed relationships between HD-Mobile cognitive task performance and DBS, rather than self-report TFC, due to the strong ceiling effect for TFC scores in premanifest-HD.

## Results

We assessed the feasibility of our study design by observing attrition rates within our participant groups. We experienced zero attrition over the course of the study. We also had impressive adherence as all participants completed all HD-Mobile cognitive tasks and questionnaires. However, due to a minor programming fault, we lost a small amount of day 8 cognitive data; three participants for Two-choice Discrimination ($n=2$ manifest-HD, $n=1$ healthy control), two participants for the OILL task ($n=2$ manifest-HD), and one manifest-HD participant for Speeded Tapping, which resulted in a slight reduction of sample size for some analyses. Initially, HD-Mobile sent task data to our server after participants returned to the main menu following the completion of each task. The participants mentioned above completed all tasks but did not return to the main menu on day 8 before exiting HD-Mobile and deleting the application, meaning some cognitive data were lost. HD-Mobile was updated to send data to the server immediately upon completion of tasks and questionnaires and no further data loss occurred. We next assessed the psychometric properties of the HD-Mobile cognitive tasks.

## Known-group validity

For Speeded Tapping, the manifest-HD group performed worse than the premanifest-HD and control groups (Fig. 2a). We found a large and significant main effect of group, $F(2, 66) = 16.89$, $p < 0.001$, $\eta_p^2 = 0.34$, with the manifest-HD group tapping slower than the healthy control and premanifest-HD groups, both with $p \leq 0.001$ and large effects, $d = 1.64$ and 1.03, respectively. The premanifest-HD and control groups did not differ in tapping speed ($p = 0.16$), although there was a medium effect size ($d = 0.59$). There was a main effect of day, Greenhouse–Geisser adjusted $F(1.84, 121.11) = 8.32$, $p = 0.001$, $\eta_p^2 = 0.11$, indicating an improvement in tapping speed over the study days, but this effect did not interact with group ($p = 0.35$).

For accuracy (percent correct) on Two-choice Discrimination, the group main effect was again large and significant, $F(2, 63) = 6.17$, $p = 0.004$, $\eta_p^2 = 0.16$ (Fig. 2b). The manifest-HD group performed less accurately than the healthy control group with a large effect, $p = 0.002$ and $d = 0.93$, but did not differ from the premanifest-HD group, $p = 0.14$, $d = 0.057$. Healthy control and premanifest-HD groups did not differ on accuracy, $p = 0.27$, $d = 0.55$. We again observed a main effect of day, $F(2, 126) = 5.43$, $p = 0.005$, $\eta_p^2 = 0.079$, with gradual improvement over the study, and no interaction with group ($p = 0.59$).

For OILL Object Identity, we found a large and significant main effect of group, $F(2, 64) = 9.64$, $p < 0.001$, $\eta_p^2 = 0.23$ (Fig. 2c). The manifest-HD group was less accurate than the healthy control group, $p = 0.001$, and premanifest-HD group, $p < 0.001$, which were both large effects, $d = 1.08$ and 1.22, respectively. There was no difference between the healthy control and premanifest-HD groups on this measure, $p = 0.78$. There was no effect of day or interaction with group (both $ps > 0.72$). In contrast, *Object Location* did not demonstrate a group effect, $F(2, 64) = 1.53$, $p = 0.23$, $\eta_p^2 = 0.046$ (Fig. 2d). There was no effect of day, $p = 0.063$, $\eta_p^2 = 0.040$, but the interaction with group was significant, $F(4, 128) = 2.52$, $p = 0.044$, $\eta_p^2 = 0.073$, which appeared to be driven by a decrease in the control group's score from day 4 to 8, whereas the other two groups improved.

## Test–retest reliability

HD-Mobile cognitive tasks demonstrated moderate to excellent intraclass correlations for most outcome variables for all three participant groups (ICCs = 0.60–0.96; Table 2). Three intraclass correlation coefficients did not reach acceptable reliability, including OILL Object Location for the

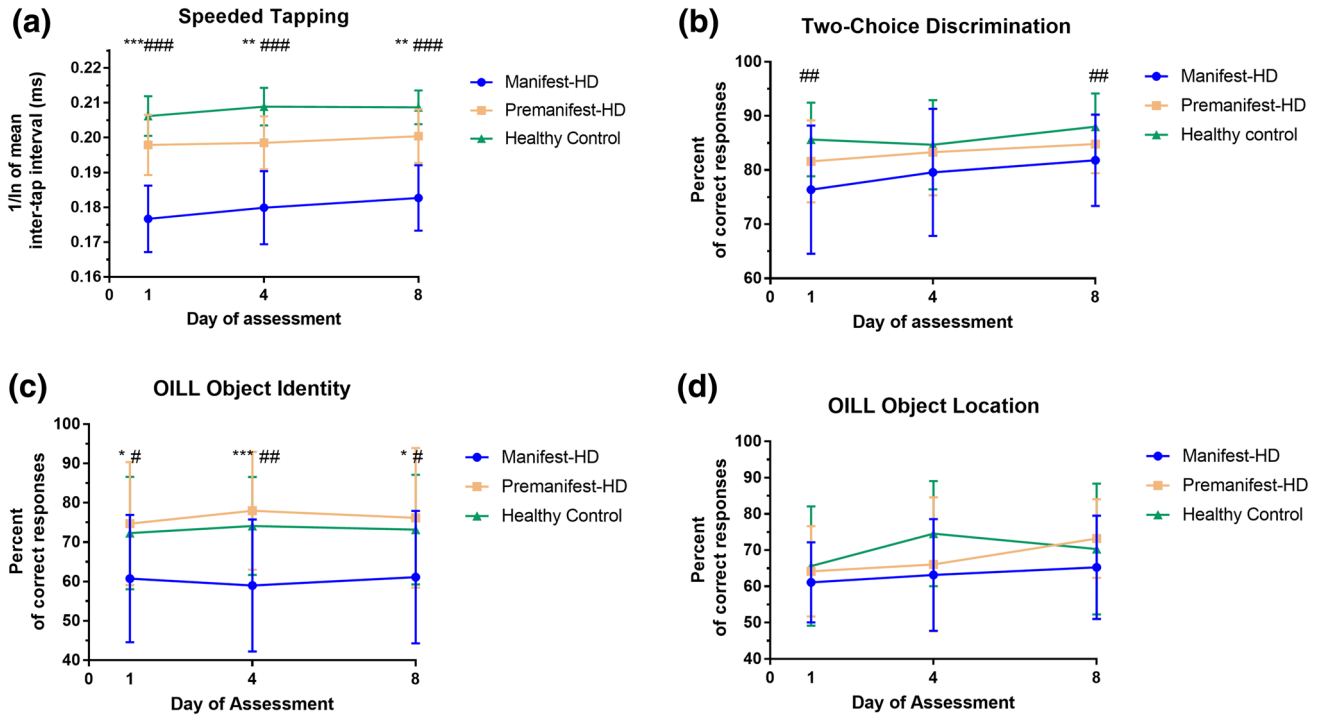



**Fig. 2** Mean and 95% confidence interval for premanifest-HD, manifest-HD, and healthy control groups performance on HD-Mobile cognitive tasks for **a** Speeded Tapping, **b** Two-choice Discrimination, **c** OILL Object Identity, **d** OILL Object Location, $^{*}p \leq 0.05$ for manifest-HD vs premanifest-HD; $^{**}p \leq 0.01$ for manifest-HD vs premanifest-HD; $^{***}p \leq 0.001$ for manifest-HD vs premanifest-HD. $^{\#}p \leq 0.05$ for manifest-HD vs healthy control, $^{\#\#}p \leq 0.01$ for manifest-HD vs healthy control; $^{\#\#\#}p \leq 0.001$ for manifest-HD vs healthy control

**Table 2** Intraclass correlation coefficients for test–retest reliability of HD-Mobile cognitive task outcome variables for premanifest-HD, manifest-HD, and control groups

| | Premanifest-HD | Manifest-HD | Healthy controls |
|---|---|---|---|
| Speeded Tapping RT | 0.96 | 0.96 | 0.96 |
| Discrimination % Correct | 0.52 | 0.77 | 0.54 |
| OILL Object Identity % Correct | 0.75 | 0.71 | 0.68 |
| OILL Object Location % Correct | 0.54 | 0.78 | 0.60 |

*Speeded Tapping RT* 1/ln inter-tap interval in milliseconds on Speeded Tapping task, *OILL Object Identity % Correct* percent correct on Object Identification and Location Learning task object identification trials, *OILL Object Location % Correct* percent correct on Object Identification and Location Learning task location identification trials, *Discrimination % Correct* percent correct on Two-choice Discrimination task

premanifest-HD group, ICC = 0.54, and percent correct on the Two-choice Discrimination for both the premanifest-HD group, ICC = 0.52, and healthy control group, ICC = 0.54 (Table 2).

## Concurrent validity

To determine concurrent validity of the HD-Mobile cognitive tasks, we assessed the relationship between performance on day 8 HD-Mobile cognitive task variables and selected Enroll-HD cognitive tasks, for the HD participants in our study who were part of Enroll-HD. For Speeded Tapping and OILL tasks, we observed medium to strong correlations ($r = 0.36$–$0.68$) with the Enroll-HD cognitive tasks, including Symbol Digits Modalities Test, F–A–S and Animals subtests from the Controlled Oral Word Association Test and the word-reading condition from the Stroop Color–Word Test, where better performance on HD-Mobile tasks correlated with better performance on Enroll-HD paper and pen tasks (Table 3). Correlations between percent correct on the Two-choice discrimination task and the Enroll-HD cognitive tasks were non-significant and at most weak, although we noted that none of the Enroll-HD tasks had comparable Two-choice decision requirements.

## Sensitivity to indicators of disease progression

We assessed the sensitivity of HD-Mobile cognitive tasks to indicators of disease progression via correlation with DBS, an indicator of age and genetically estimated disease burden, and to motor symptom progression (TMS) by analyzing correlations between day 8 cognitive task performance, DBS, and TMS across our HD participants (premanifest-HD and manifest-HD combined). Ideally, we would have preferred to use an available functional capacity measure (TFC), as we had self-report TFC data for all participants and only had TMS data for Enroll-HD participants, but ceiling effects on TFC in our sample precluded this possibility. For DBS, higher disease burden correlated with poorer performance on all HD-Mobile cognitive tasks, with $p$ values ranging from $< 0.001$ to 0.042, and effect sizes in the medium to large range, $r = -0.35$ to $-0.54$. Increased TMS had medium to strong correlations with poorer performance on OILL Object ($r = -0.57$, $p = 0.028$) and Location trials ($r = -0.71$, $p = 0.003$), but correlated non-significantly with poorer performance on Speeded Tapping ($p = 0.15$) and Two-Choice Discrimination ($p = 0.41$).

**Table 3** Pearson product-moment correlations between HD-Mobile and ENROLL-HD cognitive tasks outcome variables

| | SDMT | FAS | Animals | Stroop Word |
|---|---|---|---|---|
| Speeded Tapping RT ($n = 18$) | 0.67** | 0.48* | 0.64** | 0.47* |
| OILL Object Identity % Correct ($n = 16$) | 0.63** | 0.44 | 0.57* | 0.36 |
| OILL Object Location % Correct ($n = 16$) | 0.54* | 0.68** | 0.61* | 0.50* |
| Discrimination % Correct ($n = 16$) | −0.001 | 0.28 | 0.07 | −0.13 |

*SDMT* Symbol Digit Modalities Test, *FAS* F-A-S verbal fluency task, *Animals* animals verbal fluency task, *Stroop Word* Stroop test word trial, *Speeded Tapping RT* 1/ln inter-tap interval in milliseconds on Speeded Tapping task, *OILL Object Identity % Correct* percent correct on Object Identification and Location Learning task object identity trials, *OILL Object Location* percent correct on Object Identification and Location Learning task location identity trials, *Two-Choice Discrimination % Correct* percent correct on Two-choice Discrimination task

*$\leq 0.05$

**$\leq 0.01$

## Discussion

The key finding from this study is that a 'bring your own' approach to mobile-app-based remote assessment of cognitive and behavioral data in people with premanifest-HD and manifest-HD generates cognitive outcomes comparable to in-person testing across several aspects of psychometric validity. To our knowledge, this is the first work to conduct exclusively remote, self-administered, mobile app-based assessment of cognitive performance in a neurodegenerative population. Remarkably, we experienced no attrition in the study. All participants completed all HD-Mobile questionnaires and tasks, despite us not meeting with participants face-to-face, indicating they found the tasks, and the time commitment required to complete them, acceptable. A strength of the study is that it included both early manifest- and premanifest-HD participants, which broadens applicability to a range of levels of HD severity. We do not yet know whether the use of HD-Mobile is suitable for people with moderate to severe HD, as they were excluded from the study. Nonetheless, the findings of robust known groups and concurrent validity, sensitivity, and test–retest reliability of HD-Mobile, with some noted exceptions, indicate the app meets key recommended standards for computerized neuropsychological assessment devices set by the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology [11].

The known-group validity of HD-Mobile cognitive tasks was demonstrated as we detected the expected impaired performance of the manifest-HD group compared to controls for at least one outcome measure on all three cognitive tasks, and compared to the premanifest-HD on at least one outcome measure for two of the cognitive tasks. This shows HD-Mobile has parity with analogous in-person cognitive assessment measures [2, 23, 24]. For example, studies investigating visual memory function in HD [22, 23] observed differences between manifest-HD and controls, but not premanifest-HD and controls, which matches our findings on the Object Identity condition from the OILL task. Moreover, Speeded Tapping separated manifest-HD from premanifest-HD and controls similar to the in-person version of the task used by Stout and colleagues [2]. Only the Object Location condition in our OILL task did not separate the participant groups. Thus, overall, there appears to be limited to no loss of function or sensitivity in shifting from in-person clinician-directed assessment to remote clinician-free self-administered assessment using HD-mobile. It is possible that because our study was conducted remotely, without a researcher to provide prompts, participants may have occasionally missed the cues to remember Object Identity or Object Location, an occurrence noted by HD participants who assisted with the development of the app, resulting in them being unsure as to which information to attend. In future iterations, addition of a voice-based or on-screen prompt for each condition may improve the sensitivity and robustness of this task.

In addition to known-group validity, HD-Mobile cognitive tasks generally achieved robust concurrent validity, sensitivity, and test–retest reliability. Each is a property that new assessment tools must possess to meet basic standards of test development and to have clinical and/or scientific utility [11, 25]. In this first iteration of HD-Mobile, some analyses did not reach a priori targets for acceptable levels of reliability, sensitivity, or validity. Possible reasons for these outcomes are discussed below. We observed concurrent validity as HD-Mobile and Enroll-HD cognitive tasks measuring similar cognitive domains had medium to strong correlations in the expected directions. In contrast, Two-choice Discrimination performance was unrelated to Enroll-HD measures, probably because none of the Enroll-HD tasks assess cognitive domains relevant to the Two-choice task. A limitation of the study was our relatively small sample size. We did not apply a Bonferroni correction to correlations as this would have been overly conservative due to our small sample size and the number of comparisons. HD-Mobile cognitive tasks were sensitive to disease burden and motor symptom progression, as increased disease burden and motor symptoms generally correlated to a medium to strong degree with poorer HD-Mobile cognitive task performance. Only correlations between Two-choice Discrimination and Speeded Tapping performance with TMS did not reach target correlations. We expect at least Speeded Taping should have correlated more strongly with TMS due to the motor component of the task. Perhaps the smaller sample size for this correlation ($n = 18$) limited the ability to capture a true relationship. This can be reassessed in future studies which will have much larger sample sizes. Thus, performance on HD-Mobile cognitive tasks mostly agreed with established in-person pen-and-paper assessments and poorer performance was related to higher levels of disease burden and motor symptoms. Together, these findings provide initial evidence for the robustness and utility of HD-Mobile and further shows parity with current in-person assessment approaches.

Test–retest reliability was acceptable across the groups with a few exceptions, OILL Object Location and Two-choice Discrimination in the premanifest-HD, and Two-choice Discrimination in healthy controls. In these cases, we noted that practice effects, which are ubiquitous in cognitive testing, may have played a role. Consistent with this, we observed significant improvements across repeated testing on Two-Choice Discrimination and near significant improvement on OILL Object Location. Future

studies could be designed to distinguish between practice effects and intra-individual variability arising from natural variations in effort, mood, fatigue, etc., which reduce test–retest reliability [37]. Detecting and understanding reasons for fluctuations in day-to-day cognitive functioning is important for advising patients on how to best manage their symptoms.

Future studies can also focus on longitudinal validation of HD-Mobile on a time-scale relevant to disease progression to ensure HD-Mobile can detect within-person disease progression. The use of a remote assessment approach, using HD-Mobile, will be beneficial for measuring intra-individual variability and longitudinal change due to the ability to conduct more frequent assessments in larger samples, and at lower costs, compared to clinic-based assessments. Larger future studies with HD-Mobile could also enable an investigation of the effects of mood and sleep on cognitive function using self-report outcomes obtained within the app and will enable additional validation of the current HD-Mobile cognitive tasks and assessment of the feasibility of adding additional tasks to the app. While the addition of more tasks would provide a broader assessment of functioning, the additional time commitments may result in reductions in compliance. Future studies can aim to identify the ideal balance between comprehensiveness of assessment and participant burden. Keeping in mind that we ensured feasibility through HD-community input and a pilot study, a limitation of this study is that we do not know how many participants chose not to participate due to not being comfortable completing tasks on their phone or due to not possessing a smartphone that met the technical requirements of HD-Mobile. Future studies should aim to determine the proportion and characteristics of HD gene-positive persons who are unable or unwilling to participate in remote assessment so that researchers can carefully define the limits of generalizability of results from remote assessment studies.

In summary, our findings demonstrate that HD-Mobile performance-based cognitive tasks have promising utility for assessing cognitive functioning of HD samples in a disease stage-specific manner. More broadly, we demonstrate that smartphone-based assessment allows convenient, frequent, and potentially ongoing assessment of HD samples without the limitation of geographical barriers, allowing collection of data from more representative samples than is possible when participants must present in person at testing sites. These results justify deploying HD-Mobile in studies with significantly larger samples than this feasibility study. Our findings also have important implications for measuring cognitive functioning in HD in relation to assessing responses to therapeutics and may be particularly helpful in post-marketing surveillance of ongoing use of new therapeutic drugs in conjunction with traditional assessment methods.

**Author contributions** BM: (1) research project conception, organization, execution, recruitment of participants. (2) Statistical analysis design and execution. (3) Manuscript writing of first and subsequent drafts. SCA: (1) research project conception. (2) Statistical analysis review and critique. (3) Manuscript review and critique. YG-J: (2) statistical analysis review and critique. (3) Manuscript review and critique. E-CM: research project organization, recruitment of participants. (3) Manuscript review and critique. NWGM: research, recruitment of participants. (3) Manuscript review and critique. CL: (2) statistical analysis review and critique. (3) Manuscript review and critique. MAB: (2) statistical analysis review and critique. (3) Manuscript review and critique. JCS: (1) research project conception. (2) Statistical analysis review and critique. (3) Manuscript review and critique.

**Funding** This study was partially funded by an Early Diagnosis and Prevention grant from the Australian Dementia Collaborative Research Centers (Project number S1.16.05). Brendan McLaren: None. Sophie Andrews was supported by a fellowship from the Huntington's Disease Society of America. Yifat Glikmann-Johnston is the recipient of a National Health and Medical Council-Australian Research Council Dementia Research Development Fellowship (Grant number 1100862) funded by the Australian Government. Emily-Clare Mercieca: None. Nicholas W. G. Murray: None. Clement Loy is supported by an Australian NHMRC Dementia Fellowship (APP1107657). Mark Bellgrove (1) Mapping Cognitive Impulsivity through online testing, CI-B, Awarded by Australian Research Council (ARC), ARC, Monash University, Eastern health: 15/12/15 → 15/06/19. (2) CogChip: development of a targeted genotyping chip for execution function, CI-A, Awarded by Australian Research Council (ARC), ARC, Monash University, Illumina, Hudson Institute of Medical Research: 20/12/17 → 19/12/20. (3) A dimensional approach to mapping the risk mechanisms of mental illness, CI-B, Awarded by National Health and Medical Research Council (NHMRC), NHMRC: 20/12/17 → 19/12/20. (4) Cellular Modelling of Attention Deficit Hyperactivity Disorder risk genes, CI-C, Awarded by National Health and Medical Research Council (NHMRC), NHMRC: 20/12/17 → 19/12/20. (5) Mechanisms and contexts driving Impulsivity, CI-C, Awarded by Australian Research Council (ARC), ARC, Monash University: 1/01/18 → 31/12/20. (6) Neurophysiology of attention deficits after right hemisphere stroke, CI-A, Awarded by National Health and Medical Research Council (NHMRC), NHMRC: 1/01/18 → 31/12/20. (7) The influence of attentional selection on perceptual decision making, CI-A, Awarded by Australian Research Council (ARC), ARC, Monash University, Trinity College Dublin: 1/01/18 → 31/12/20. (8) Neurobiology of decision making, CI-A, Awarded by the Office of Naval Research—Global, Office of Naval Research—Global: 22/08/18 → 21/09/20. (9) When Autism and ADHD overlap: finding biologically defined subtypes across the spectrum, CI-B, Awarded by the Avant Foundation, Avant Foundation, 17/10/18 → 17/10/19. (10) Systems Neuroscience of Attention Deficit Hyperactivity Disorder (ADHD), CI-A, Awarded by National Health & Medical Research Council (NHMRC), NHMRC: 1/01/19 → 31/12/23. (11) Data-driven diagnoses and treatments for neurodevelopmental disorders, CI-A, Awarded by the Australian Government Department of Health, Australian Government Department of Health: 20/01/20 → 30/06/24. (12) Utilising digital training technology to support cognitive skill development and enhance, learning in primary school children, CI-C, Awarded by the Ian Potter Foundation and Monash University, Ian Potter Foundation, Monash University: 20/01/20 → 30/06/24. (13) Understanding ADHD in Indigenous Children using the longitudinal study of Indigenous Children (LSIC), CI-A, Awarded by the Australian ADHD Professionals Association (AADPA), AADPA, 7/02/20 → 30/12/20. Julie Stout: Disclosure as of March 4, 2020: Clinical Study: Implementation of HD-CAB, Vaccinex Inc.: Signal-HD, Contract to Stout Neuropsych Pty Ltd, with subcontract to Monash University, total value approximately $520,777 AUD,

May 2015—April 2020. Consulting agreement, uniQure, Value not specified, estimated at less than $10: per annum, May 2018 to present. Consulting agreement, CHDI Foundation, Inc. Value up to $60K, June 2018 to present. Consulting agreement, Sage Therapeutics, Value up to $23,000, August 2019 to present. Since 2014, Dr. Stout has been director of a CRO that implements cognitive assessments in clinical trials (Stout Neuropsych), which has included the activities disclosed above for the following companies: Teva, Omeros, Vaccinex, and Ionis. Existing contracts go through 2021, and through this company Dr. Stout indirectly (through a family trust) receives remuneration. Dr Stout is also director of Zindametrix, a CRO that provides research support to uniQure.

**Data availability** Study data can be made available upon reasonable request.

**Code availability** SPSS syntax used for analyses can be made available upon reasonable request.

## Compliance with ethical standards

**Conflicts of interest** Since 2014, Julie C. Stout has been the director of a CRO that implements cognitive assessments in clinical trials (Stout Neuropsych), which has included the activities disclosed above for the following companies: Teva, Omeros, Vaccinex, and Ionis. Existing contracts go through 2021, and through this company, Dr. Stout indirectly (through a family trust) receives remuneration. Dr Stout is also the director of Zindametrix, a CRO that provides research support to uniQure.

**Ethics approval** This study was approved by the Monash University Human Research Ethics Committee (project number CF16/280–2016000126), the Calvary Healthcare Bethlehem Research Ethics and Ethics Committee (project number 1608103), and the Western Sydney Local Health District Human Research Ethics Committee (project number HREC/18/WMEAD/189). This study was granted ethical approval by three appropriate ethics committees. The research was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments. All persons gave their informed consent prior to their inclusion in the study and no personally identifiable data has been presented in this manuscript.

**Consent to participate** All participants gave informed consent to participate. The study was conducted under ethical approval of the above-listed institutions and adhered to the guidelines of the 1964 Declaration of Helsinki and its later amendments. All persons gave their informed consent prior to their inclusion in the study and no personally identifiable data has been presented in this manuscript..

**Consent for publication** All participants gave consent for their study data to be published in a de-identified format.

## References


1. Maroof DA, Gross AL, Brandt J (2011) Modeling longitudinal change in motor and cognitive processing speed in presymptomatic Huntington's disease. J Clin Exp Neuropsychol 33(8):901–909. https://doi.org/10.1080/13803395.2011.574606
2. Stout JC, Queller S, Baker KN, Cowlishaw S, Sampaio C, Fitzer-Attas C, Borowsky B, Barch D, Sahakian B, Bilder R, Salmon D, Brunner D, Giuliano J, Vaccarino A, Howland D, van Kammen D, Keefe R, Warner J, Marder K, Duff K, Sutherland M, Thompson J (2014) HD-CAB: a cognitive assessment battery for clinical trials in Huntington's disease. Mov Disord 29(10):1281–1288. https://doi.org/10.1002/mds.25964
3. Cardoso F (2009) Huntington disease and other choreas. Neurol Clin 27(3):719–736. https://doi.org/10.1016/j.ncl.2009.04.001
4. Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N (2012) The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. Mov Disord 27(9):1083–1091. https://doi.org/10.1002/mds.25075
5. Eddy CM, Rickards HE (2013) Impact of cognitive and behavioural changes on quality of life in Huntington's disease. Basal Ganglia 3(2):123–126. https://doi.org/10.1016/j.baga.2013.01.085
6. Ross CA, Pantelyat A, Kogan J, Brandt J (2014) Determinants of functional disability in Huntington's disease: role of cognitive and motor dysfunction. Mov Disord 29(11):1351–1358
7. Rodrigues FB, Quinn L, Wild EJ (2019) Huntington's disease clinical trials corner: January 2019. J Huntington's Dis 8(1):115–125
8. Long JD, Lee J-M, Aylward EH, Gillis T, Mysore JS, Elneel KA, Chao MJ, Paulsen JS, MacDonald ME, Gusella JF (2018) Genetic modification of Huntington disease acts early in the prediagnosis phase. Am J Hum Genet 103(3):349–357
9. Moss DJH, Pardiñas AF, Langbehn D, Lo K, Leavitt BR, Roos R, Durr A, Mead S, Coleman A, Santos RD (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol 16(9):701–711
10. Moore RC, Swendsen J, Depp CA (2017) Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 26(4):e1562. https://doi.org/10.1002/mpr.1562
11. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI (2012) Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology(). Arch Clin Neuropsychol 27(3):362–373. https://doi.org/10.1093/arclin/acs027
12. Dufau S, Duñabeitia JA, Moret-Tatay C, McGonigal A, Peeters D, Alario FX, Balota DA, Brysbaert M, Carreiras M, Ferrand L, Ktori M, Perea M, Rastle K, Sasburg O, Yap MJ, Ziegler JC, Grainger J (2011) Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. PLoS ONE. https://doi.org/10.1371/journal.pone.0024974
13. Fredrickson J, Maruff P, Woodward M, Moore L, Fredrickson A, Sach J, Darby D (2010) Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. Neuroepidemiology 34(2):65–75. https://doi.org/10.1159/000264823
14. Falleti MG, Maruff P, Collie A, Darby DG (2006) Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. J Clin Exp Neuropsychol 28(7):1095–1112
15. Chinner A, Blane J, Lancaster C, Hinds C, Koychev I (2018) Digital technologies for the assessment of cognition: a clinical review. Evid Based Mental Health 21(2):67. https://doi.org/10.1136/eb-2018-102890
16. Daniëls NEM, Bartels SL, Verhagen SJW, Van Knippenberg RJM, De Vugt ME, Delespaul PAEG (2020) Digital assessment of working memory and processing speed in everyday life: feasibility, validation, and lessons-learned. Internet Interv 19:100300. https://doi.org/10.1016/j.invent.2019.100300
17. Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E (2017) Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. JMIR Mhealth Uhealth 5(5):e68. https://doi.org/10.2196/mhealth.6939
18. Tabrizi SJ, Reilmann R, Roos RAC, Durr A, Leavitt B, Owen G, Jones R, Johnson H, Craufurd D, Hicks SL, Kennard C,

Landwehrmeyer B, Stout JC, Borowsky B, Scahill RI, Frost C, Langbehn DR (2012) Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. Lancet Neurol 11(1):42–53. https://doi.org/10.1016/S1474-4422(11)70263-0
19. Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput 20(4):873–922
20. Herz DM, Little S, Pedrosa DJ, Tinkhauser G, Cheeran B, Foltynie T, Bogacz R, Brown P (2018) Mechanisms underlying decision-making as revealed by deep-brain stimulation in patients with Parkinson's disease. Curr Biol 28(8):1169–1178.e1166. https://doi.org/10.1016/j.cub.2018.02.057
21. Huang Y-T, Georgiev D, Foltynie T, Limousin P, Speekenbrink M, Jahanshahi M (2015) Different effects of dopaminergic medication on perceptual decision-making in Parkinson's disease as a function of task difficulty and speed–accuracy instructions. Neuropsychologia 75:577–587
22. Glikmann-Johnston Y, Halag-Milo T, Hendler T, Gazit T (2020) Functional MRI (fMRI) human mapping for indicating hippocampal content-specific responsivity and laterality. Neuropsychology 34:578–590
23. Begeti F, Schwab LC, Mason SL, Barker RA (2016) Hippocampal dysfunction defines disease onset in Huntington's disease. J Neurol Neurosurg Psychiatry 87(9):975–981
24. Glikmann-Johnston Y, Carmichael AM, Mercieca E-C, Stout JC (2019) 'Real-life'hippocampal-dependent spatial memory impairments in Huntington's disease. Cortex 119:46–60
25. American Psychological Association (2017) Ethical principles of psychologists and code of conduct. https://www.apa.org/ethics/code/. Accessed 02 Feb 2020
26. Penney JB, Vonsattel JP, Macdonald ME, Gusella JF, Myers RH (1997) CAG repeat number governs the development rate of pathology in Huntington's disease. Ann Neurol 41(5):689–692
27. Zhang Y, Long JD, Mills JA et al (2011) Indexing disease progression at study entry with individuals at-risk for Huntington disease. Am J Med Genet Part B Neuropsychiatr Genet 156B(7):751–763. https://doi.org/10.1002/ajmg.b.31232
28. Buffalo EA, Bellgowan PS, Martin A (2006) Distinct roles for medial temporal lobe structures in memory for objects and their locations. Learn Mem 13(5):638–643
29. Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67(6):361–370
30. Cosco TD, Doyle F, Ward M, McGee H (2012) Latent structure of the hospital anxiety and depression scale: a 10-year systematic review. J Psychosom Res 72(3):180–184
31. De Souza J, Jones LA, Rickards H (2010) Validation of self-report depression rating scales in Huntington's disease. Mov Disord 25(1):91–96
32. (1996) Unified Huntington's disease rating scale: reliability and consistency. Mov Disord 11 (2):136–142. https://doi.org/10.1002/mds.870110204
33. Carlozzi NE, Boileau NR, Perlmutter JS, Chou KL, Stout JC, Paulsen JS, McCormack MK, Cella D, Nance MA, Lai J-S (2018) Agreement between clinician-rated versus patient-reported outcomes in Huntington disease. J Neurol 265(6):1443–1453
34. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, Cheng WY, Fernandez-Garcia I, Siebourg-Polster J, Jin L (2018) Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 33(8):1287–1297
35. Walker FO (2007) Huntington's disease. Lancet 369(9557):218–228
36. Broglio SP, Ferrara MS, Macciocchi SN, Baumgartner TA, Elliott R (2007) Test-retest reliability of computerized concussion assessment programs. J Athl Train 42(4):509–514
37. Strauss E, Sherman EM, Spreen O (2006) A compendium of neuropsychological tests: administration, norms, and commentary. American Chemical Society, Washington

# Chapter Four: Study Three

## Diffusion Modeling of a Two-choice Decision-making Task in Premanifest and Manifest Huntington's Disease

### 4.1. Explanatory Note

In this chapter, I present Study Three from the thesis research. In this chapter I present further work which focused on addressing our second research aim, which was to demonstrate that the cognitive tasks we developed for the research project would be reliable, and valid, across several forms of validity, and sensitive to cognitive impairment in HD compared to controls. The study presented in Chapter Four provided an important extension of the work presented in Chapter Three. In Chapter Three, I demonstrated that HD-Mobile cognitive tasks had discriminant validity in that they could separate the manifest-HD group from the premanifest-HD and healthy control groups based on cognitive performance. The analysis strategy that we used in Chapter Three, however, did not detect cognitive performance differences between the premanifest-HD and healthy control groups. In chapter Four, we applied cognitive modelling to data from the Two-choice Discrimination task, which enabled us to detect cognitive performance deficits in premanifest HD versus healthy control groups. The results from Study Three provided further demonstration of the discriminant validity of HD-mobile and also provided a further demonstration that sensitive and informative cognitive assessment can be achieved using smartphones and without ever needing to meet patients or participants in person.

**Note.** This chapter was submitted to the *Journal of the International Neuropsychological Society* and is presented in its submitted format. Within this chapter, there are several references to supplementary materials. The supplementary materials that were submitted with this paper are presented in Appendix 3.

Diffusion Modeling of a Two-choice Decision-making Task in Premanifest and Manifest Huntington's Disease

Brendan McLaren[1], BBioSc (Hons), GradDipPsych, BSc (Hons), ORCID iD 0000-0002-8769-8629, Ricardo J. Romeu[2], BA, Sophie C. Andrews[1,,3,4], BA, GradDipPsych, BSc (Hons), DPsych (Clin Neuro), ORCID iD 0000-0002-2103-7748, Mark A. Bellgrove[1], BSc (Hons), PhD, ORCID iD 0000-0003-0186-8349, Julie C Stout[1], BSc, MA, PhD[1], ORCID iD 0000-0001-9637-8590

[1]*Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, Victoria, Australia*

[2]*Cognitive Science Program, Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, United States of America*

[3]*Neuroscience Research Australia, Sydney, NSW, Australia*

[4]*School of Psychology, University of New South Wales, Sydney, Australia*

**Correspondence to**: Prof. Julie C Stout, Turner Institute for Brain and Mental Health 18 Innovation Walk, School of Psychological Sciences, Monash University, Clayton, Vic 3800, Australia; Email: Julie.stout@monash.edu; Phone: +61 3 9905 3987

Word count

Abstract: 250

Manuscript: 4981

## Abstract

**Objective:** Sensitive cognitive assessment is essential for monitoring cognitive change and response to disease modifying interventions in Huntington's disease (HD). We applied a hierarchical Bayesian diffusion model to decision-making data collected from HD-Mobile a smartphone-based cognitive assessment application. We aimed to determine the model's fitness for interpretation and whether the model could detect decision-making performance deficits between people with premanifest-HD and manifest-HD compared to healthy control participants.

**Method** Forty-two HD CAG-expanded participants (20 manifest, 22 premanifest) and 29 healthy controls completed HD-Mobile cognitive tasks three times across an eight-day period, on days one, four, and eight, along with several self-report questionnaires. Participants completed all tasks independently in their homes. We applied a hierarchical Bayesian model to data from the Two-Choice Discrimination decision task and examined Pearson's correlations between model outcomes, cognitive data from the other HD-Mobile tasks, and questionnaire data.

**Results** Model data was sufficiently reliable to allow group-level comparisons across all four key parameters produced by the model, and to enable correlation analysis of individual participant performance for two of the model parameters with disease severity variables. The model was highly sensitive, detecting performance differences across groups, with both manifest- and premanifest-HD performing worse than healthy controls. We also detected several significant correlations between model parameters and self-reported measures of sleep and sleepiness.

**Conclusions** Diffusion modeling using hierarchical Bayesian analysis of HD-Mobile cognitive task data is highly sensitive for detecting cognitive signs in both premanifest- and

manifest-HD, outperforming behavioral measures from the same task in detecting differences between the premanifest-HD and control groups.

*Keywords*: Huntington's Disease, Diffusion Modeling, Bayesian Analysis, Sleep Quality, Clinical Mathematical Psychology, cognition, neurodegeneration

# Diffusion Modeling of a Two-choice Decision-making Task in Premanifest and Manifest Huntington's Disease

Huntington's disease (HD) is an autosomal dominant, neurodegenerative disease, which has a typical onset in middle adulthood, and results in progressive cognitive, motor, and psychiatric symptoms (Walker, 2007). The cognitive symptoms in HD are often detectible many years before clinical diagnosis (Maroof, Gross, & Brandt, 2011; Stout et al., 2014), contribute significantly to declines in functioning and quality of life (Eddy & Rickards, 2013, 2015; Ross et al., 2014), and result from widespread neuronal atrophy which is most severe in the striatum (Walker, 2007). Striatal atrophy and accompanying breakdown of fronto-striatal networks are associated with cognitive deficits (e.g., Bamford, Caine, Kido, Cox, & Shoulson, 1995; Campodonico et al., 1998; Starkstein et al., 1992) and behavioral changes such as impulsivity and impaired judgment (Johnson, Potts, Sanchez-Ramos, & Cimino, 2017; Mörkl et al., 2016).

Large longitudinal studies of HD, typically utilizing annual cognitive testing, have contributed to what is now a strong account of how the cognitive symptoms of HD develop and evolve over time (e.g., Paulsen, Smith, & Long, 2013; Tabrizi et al., 2013). Recently, our group has developed smartphone-based mobile apps to enable us to shift some cognitive assessments to be self-administered independently, facilitating testing in more ecologically-relevant settings (i.e., home) and at intervals that enable frequent sampling of cognition and the possibility of understanding the dynamics of cognitive function in relation to regular elements of daily life, such as sleep and physical activity (McLaren et al., 2020).

In this study, we extended this at-home assessment approach one step further by applying formal cognitive modeling to data collected on the mobile app, further extending the possible utility and value of home-based, unsupervised data collection approaches. Cognitive

modelling has been applied only rarely in neurodegenerative diseases (Busemeyer & Stout, 2002; Herz et al., 2018; Huang et al., 2015) but not to data collected from smartphones. This approach opens the potential to move beyond observable measures such as errors and response time to examine and quantify the internal relationships within data series using equations that can reveal and quantify the interplay of cognitive processes that yield the observable outcomes (Busemeyer & Diederich, 2010). To our knowledge, no previous studies have attempted to apply cognitive models to datasets collected by participants with neurodegenerative diseases in an unsupervised setting. If such approaches turn out to be fruitful, however, this approach raises the possibility of building a rich understanding of cognitive impairment using much larger scale and lower cost data collections, in conjunction with a host of other measures such as passive activity monitoring, which markedly adds to the potential for understanding of these diseases (Allard et al., 2014; Kaye et al., 2011; McCrae, Vatthauer, Dzierzewski, & Marsiske, 2012) .

For our study, to facilitate self-administration in people with cognitive impairment, we selected a simple two-choice decision-making task (McLaren et al., 2020). The task we used has homologies to a classical neuropsychological task, the Wisconsin Card Sorting Test, in that participants are provided with little instruction, and must make a series of selections, each followed by feedback, enabling them to learn by trial and error to improve choice accuracy. In our task (Figure 1), participants observed a scene with snowflakes falling from the sky and chose either the “few” or “many” response option. Each choice was followed by feedback “correct” or “incorrect.” Three hundred and sixty trials were completed, requiring about 10 – 15 minutes, providing the opportunity for learning from experience to improve choice accuracy.

We applied the classic cognitive model for two-choice decision task, the Ratcliff Diffusion Model (RDM; Ratcliff, 1978; Ratcliff & McKoon, 2008; White, Ratcliff, Vasey, &

McKoon, 2010), to the data from this study. The RDM assumes that stimuli related to a decision process are translated into "evidence," accumulated randomly until one of two decision boundaries (i.e., "few," "many") are reached (see Figure 2). The evidence accumulation process unfolds over time, making it possible to extract response time predictions from the model by characterizing response latencies (i.e., latency to reach a decision boundary). In the RDM, the decision process is modeled as several components, or parameters, each instantiated in a term in the model equation, from which values can be derived for each participant (i.e., parameter estimates). In effect, this yields additional model-generated dependent measures that can be studied along with other study outcomes to provide information about a person's cognitive functioning. Specifically, the four parameters derived from the RDM are detailed in Table 1. Henceforth, to facilitate readability, in this paper RDM variables will be referred to by their psychological meanings, as per Table 1. For example, we will use "evidence accumulation speed" in place of the parameter term Drift Rate.

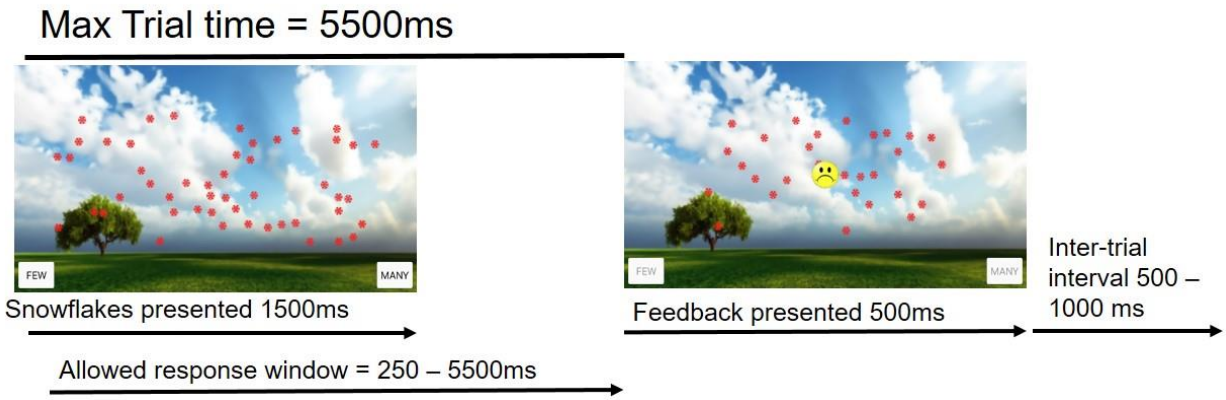



Figure 1. Schematic diagram of the Two-choice Discrimination task. Trial durations were 250 – 5500 milliseconds (ms). Snowflakes were presented for up to 1500ms, then disappeared from the screen to prevent counting. Participants had 5500 ms to respond. Responses ranging from 250 to 5500 ms triggered feedback (i.e., happy or sad emoticon) for 500 ms followed by a 500 ms inter-trial interval. A 1000 ms inter-trial interval occurred if responses were out of window. Response buttons were disabled during feedback and the inter-trial interval.

An additional key process in applying cognitive modeling is the application of formal quality checks to the model-generated outcomes. This step evaluates whether the model fits are able to account for the data well enough to engender confidence in the subsequent

interpretation of the resulting model outcomes. The adequacy of model fits may be jeopardized in cases of insufficient or noisy data, yielding parameter estimates that are unacceptably biased, resulting in data that is unreliable and unfit for interpretation (Kruschke, 2014). Our study was clearly at risk for problems with data quality given that we used a novel, self-administration testing approach in cognitively impaired people, and we were only able to access in a reasonably limited sample size. To address this concern, we applied the RDM using a Bayesian analysis, which is better equipped to handle sparse and noisier data (Kruschke, 2014; Wagenmakers et al., 2018), and subsequently limited our focus to the modeling results that were sufficiently robust. This method also facilitated our understanding of the limitations of the dataset we generated, for studies incorporating cognitive modeling.

Thus, our aims in this study were to determine: 1) whether we could use data collected by self-administration in people with the HD CAG expansion in the context of cognitive modeling, to yield parameter estimates that met the thresholds for interpretability; 2) whether new variables generated from the model, which provide an account of the cognitive operations at play in the decision-making process, would demonstrate known groups validity. That is, they would be able to distinguish between the HD and healthy control groups; and, 3) whether the parameter estimates would relate to disease severity or other measures of participant function, including other cognitive tasks collected by self-administration, and self-reports of sleep and alertness. We expected that the parameter estimates, at least in the evidence accumulation speed (Drift Rate) and perhaps encoding/motor execution time (Non-decision Time), would relate to disease severity. Specifically, we hypothesized that evidence accumulation speed would be lowest in the manifest-HD group and greatest in the control group, reflecting slowed processing speed, with the premanifest-HD group midway between manifest-HD and controls.

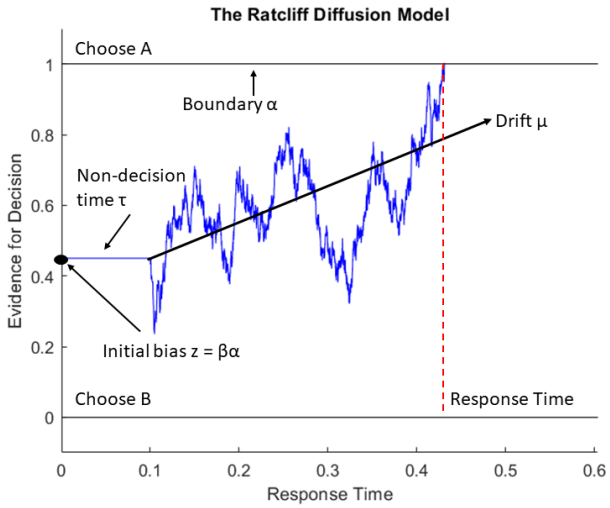



*Figure 2.* A Sample Path of the Ratcliff Diffusion Model. The graph is a simulation of a single trial, wherein a participant needs to decide between options A and B; option A is arbitrarily assigned to the top boundary (which in this example happens to be one). The drift represents the average direction of the sample path over time; since it is aimed towards the upper boundary, this drift rate is positive. The bias is the value of the sample path at time zero – it is where the sample path begins. Note that in this implementation of the model, the bias is obtained as a proportion β of the boundary α. The length of the horizontal line before the diffusion path represents the non-decision time. The response time is recorded as the time the path hits a boundary, plus the non-decision time.

Table 1.

*Ratcliff Diffusion Model Parameters*

| **Concept** | **Symbol** | **Effect in the Model** | **Psychological Meaning** | **Possible Neural Pathway** |
|---|---|---|---|---|
| Drift | μ | Slope of general direction of diffusion | Measures rate and direction of general evidence accumulation/encoding speed | Fronto-Parietal Pathway |
| Boundary | α | When the process hits this boundary, a particular decision is made | A measure of how cautious the person is in the decision task; how much evidence is needed until a decision can be made | Fronto-Basal Ganglia Pathway |
| Bias | β | Starting point, relative to boundary, of | Measuring response bias in the person across trials | Both Fronto-Parietal and Fronto-Basal |

| | | | | |
|---|---|---|---|---|
| | | diffusion process at the beginning of a trial | | Ganglia Pathways |
| Non-decision Time | τ | Recorded Response time broken up into Process Time + Non-Decision Time | Typically interpreted as encoding time and motor/execution time, neither of which is otherwise measured in the standard model | Unknown – Too vague a concept to map to neural functioning |
| Variability in Drift* | σ | Spread of mean drift rates within a group | A measure of the amount of individual difference in evidence accumulation rate within a group | N/A – A measure between people |

*Note.* A description of the major parameters involved in the Ratcliff Diffusion Model, including what they do in the model and how they are typically interpreted cognitively. The pathway suggestions are based on a review done by (Mulder et al., 2014) that correlates BOLD signals with drift diffusion parameter estimates. *Variability in drift is not typically included in the model, but our implementation found this to be an important factor. See Supplement below. It may help the reader to compare this Table with Figure 2.

## Method

### Participants

Our sample included 22 participants with premanifest HD, 20 with early manifest HD, and 29 healthy controls (Table 2). Participants were required to be 18 – 70 years of age, fluent in English, and to have an Android or Apple smartphone running Android 4.0 (or above) or iOS 7 (or above) operating systems. We excluded participants with a history of neurological illness (other than HD), significant head injury, psychiatric disorder other than depression or anxiety, drug abuse, or current involvement in any clinical drug trial. Manifest-HD participants required prior diagnosis by an HD specialist clinician. Premanifest participants were required to have a CAG expansion of at least 41 and disease burden score (DBS) of ≥ 200 (calculated as age x [CAG – 35.5]; Penney, Vonsattel, Macdonald, Gusella, & Myers, 1997). We also calculated the CAG-Age product for premanifest- and manifest-HD participants (calculated as age x [CAG – 33.6]; Zhang et al., 2011). The manifest-HD group

was significantly older than the premanifest HD and healthy control groups ($p$ = .005 and .027, respectively) and had lower self-reported Total Functional Capacity (Unified Huntington's Disease Research Rating Scale [UHDRS], Huntington Study Group, 1996) than premanifest HD and healthy control groups (both $ps$ < .001), as expected. The manifest-HD group had significantly higher total motor score (TMS; (UHDRS, Huntington Study Group, 1996), DBS, and CAG-Age product scores than the premanifest group (all $ps$ < .008). The premanifest-HD group had less education than the healthy control group ($p$ = .012). There were no other significant differences between groups (all $p$s > .10).

We recruited participants through databases at Monash University, Melbourne, and Westmead Hospital, Sydney, which include volunteers who have consented to being contacted for research opportunities. Additional participants were recruited using advertisements for the study posted on Facebook and at local HD community events.

Table 2.
*Participant Characteristics*

| Characteristic | Healthy Controls | Premanifest-HD | Manifest-HD |
|---|---|---|---|
| *n* | 29 | 22 | 20 |
| Age (Mean Years (SD), Range) | 44.31 (10.93)<br>28 - 65 | 42.09 (8.52)<br>21 - 58 | 51.60 (8.03)[a, c]<br>31 - 65 |
| Education (Mean Years (SD), Range) | 14.88 (1.68)<br>12 -19 | 13.25 (2.09)[c]<br>10 - 18 | 14.33 (2.14)<br>12 - 18 |
| CAG Repeat length (Mean Years (DS), Range) | - | 42.95 (2.16)<br>41 - 51 | 42.88 (2.11)<br>41 - 49 |
| Disease-burden score (Mean (SD), Range) | - | 304.12 (56.13)<br>217.50 – 435.00 | 379.64 (58.85)<br>275.00 – 484.50 |
| CAG-Age Product (Mean (SD), Range) | | 382.47 (66.45)<br>270.86 – 541.72 | 475.62 (60.78)<br>367.00 – 589.38 |
| HADS Anxiety (Mean (SD), Range) | 6.00 (4.53)<br>0 – 16 | 6.36 (4.29)<br>0 – 15 | 5.85 (3.62)<br>2 – 17 |
| HADS Depression (Mean (SD), Range) | 3.17 (3.42)<br>0 – 16 | 3.50 (3.26)<br>0 – 11 | 5.40 (4.45)<br>0 – 14 |
| UHDRS TFC (Self-report) (Mean (SD), Range) | 12.97 (0.19)<br>12 - 13 | 12.64 (0.90)<br>9 - 13 | 9.26 (2.42)[b, d]<br>6 - 13 |
| *n* | - | 12 | 7 |

*Note.* HADS = Hospital Anxiety and Depression Scale, UHDRS TFC = Unified Huntington's Disease Rating Scale Total Functional Capacity, UHDRS TMS = Unified Huntington's Disease Rating Scale Total Motor Score, which was only available for participants involved with Enroll-HD, [a] $p$ < .05 with respect to premanifest-HD, [b] $p$ < .01 with respect to premanifest-HD, [c] $p$ < .05 with respect to controls, [d] $p$ < .01 with respect to controls.

**Materials**

We developed a custom designed smartphone application, HD-Mobile, for this study. The application was designed for touch screen smartphones operating Android 4.0 and iOS 7 and above. HD-Mobile was used to send automated reminder messages to participants, to administer cognitive tests and self-report questionnaires, and to automatically transmit data to the lab server. HD-Mobile included the Two-choice Discrimination task, as modeled after the Numerosity Discrimination task described by Ratcliff and McKoon (2008). Participants viewed a scene depicting falling snowflakes and were instructed to quickly decide, without counting, whether there were "few" or "many" snowflakes on the screen. To facilitate learning, they received immediate feedback after each choice, either a happy (correct choices) or sad face (incorrect choices). Three blocks of 40 trials were presented at each time-point (days one, four, and eight), (see Figure 2). The cut-off value for each block, above which the "many" response was correct and below which the "few" response was correct, was randomized between the values of 45 and 55 snowflakes. We randomized the number of snowflakes per trial to ± 20 of the cut-off value, maintaining at least 40% of trials for each response ("few", "many"). We analyzed "few" and "many" responses as separate conditions, as the "few" condition was designed to be easier and the "many" to be more challenging.

HD-Mobile included two additional tasks, Object Information and Location Learning (OILL) and Speeded Tapping, and questionnaires, described elsewhere (McLaren et al., 2020). Briefly, OILL assesses Object Identity and Object Location working memory, recorded as percent correct responses and mean response time. Speeded Tapping is a bimanual speeded thumb-tapping task, used to assess psychomotor speed.

Self-report questionnaires included the SCale of Outcomes in PArkinsons disease – Sleep (SCOPA –SLEEP) (Marinus, Visser, van Hilten, Lammers, & Stiggelbout, 2003), the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), and the

Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990), as described in McLaren et al (2020). Briefly, SCOPA-SLEEP assesses night-time sleep (5 items, termed "Sleep Disturbance") and daytime sleepiness (6 items, termed "Daytime Sleepiness"), and includes a question about Sleep Quality in the past month. Higher subscale scores indicate poorer night-time sleep and more daytime sleepiness. The HADS is a widely used and reliable (Cosco, Doyle, Ward, & McGee, 2012; De Souza, Jones, & Rickards, 2010; Zigmond & Snaith, 1983) 14-item self-report questionnniare, which assesses feelings of anxiety and depression over the previous week. The KSS is a single question indicating subjective alertness at the present time.

**Procedure**

We provided instructions for installation of HD-Mobile and study procedures in written and video formats. Participation occurred across eight days, with at least one study task per day. Cognitive tests were on days one, four, and eight, yielding about 360 data points per person for the Two-choice Discrimination task. We requested that tasks be completed within an hour of waking on each day. Participants were messaged by HD-Mobile daily at 9am to remind them to complete their daily tasks (unless already complete). Across the eight days, the study commitment required under 2.5 hours.

**Data Analysis**

We used IBM SPSS V.26 for basic descriptive statistics, group comparisons, and correlational analyses, excluding missing data listwise. Alpha was set at .05 for all analyses. We assessed demographic group differences using one-way analysis of variance and independent measures t-tests. We used Pearson's to examine correlations between parameter estimates from modeling and the other cognitive outcomes and, demographic and clinical variables.

**Data Modeling Using Hierarchical Bayesian Diffusion Analysis**

We fit performance data to the Ratcliff Diffusion Models using a hierarchical Bayesian version of the RDM (see Vandekerckhove, Tuerlinckx, & Lee, 2011; Wiecki, Sofer, & Frank, 2013), which yields separate sets of parameter estimates, one set for each group, a second set for each participant, for use respectively in group comparisons, and in individual differences correlations (e.g., associations with clinical variables; (Kruschke, 2014; Wagenmakers et al., 2018).

To ensure the accuracy and reliability of the parameter estimates, we first tested several models to assess the explicit assumptions of the Bayesian approach. We then selected the model best for minimizing the deviance from the data, akin, in principle, to residuals in regression analysis, using the Deviance Information Criterion method (DIC; Fox, 2010; Spiegelhalter, Best, Carlin, & Van der Linde, 2014), as summarized here with full details of comparisons in the supplement. Two main classes of models were fit: (1) a *variability* model, which allowed differing variability in drift rates for each group, and (2) the *age-regression* model, wherein mean drift rate was a regression equation based on participant age. The *age-regression* model was selected because age is known to affect response times, and thus needed to be statistically controlled (Ratcliff, 2008). In the end, however, the variability model had the better combination of low DIC and high reliability in the parameter estimates, hence, we present results from the *variability model*.

We assessed reliability of the parameter estimates using the Gelman-Rubin statistic $\hat{R}$ (Gelman & Rubin, 1992). $\hat{R}$ values less than 1.1 are considered to be more reliable and therefore appropriate for interpretation. We examined differences in parameter estimates between groups using the Bayesian hypothesis testing method (detailed in the supplement); like the null hypothesis significance testing, this approach uses a pre-set level of dissimilarity (e.g., $p = .05$). Rather than "significantly different," however, the term used in Bayesian

hypothesis testing is “credibly different”, to highlight the philosophical differences in the two approaches. Another difference from null hypothesis testing is that Bayesian approach yields probability estimates of any directional differences. A probability of observed difference of 50%, suggests that the two parameters are approximately equal, while deviations from 50% (either in the direction of 0% or 100%) indicate likely differences between the groups. Hence, even if two groups are not credibly different, we can still describe how similar or dissimilar they were and the direction of any differences.

## Results

We first summarize the main reliability and model comparison results, with more details in the supplement, followed by a group-level comparison of parameter estimates. We then highlight the relationships observed through correlation between the RDM individual estimates and other task variables.

We found that all four RDM parameters, evidence accumulation speed (Drift Rate), response cautiousness (Boundary), response bias (Bias), and encoding/motor execution time (Non-Decision Time), were reliable at the *group level* (Gelman-Rubin statistic $\hat{R}$’s for all *group-level* parameters were less than 1.1). In contrast, of the *individual-level* parameters, only evidence accumulation speed and response cautiousness estimates were sufficiently reliable to include in the planned correlation analyses. The bias parameter $\hat{R}'s$ exceeded the 1.1 in 14% of cases and the encoding/motor execution time parameter $\hat{R}'s$ exceeded 1.1 in 90% of cases. Thus, whereas we were able to examine *group level* differences for all four parameters, at the *individual level*, we only interpreted evidence accumulation speed and response cautiousness parameters.

### Group Level Comparisons of Parameter Estimates

Group comparisons generally showed similarities in RDM parameters across groups (Table 3), although several credible differences in drift rates were observed. Specifically, the

manifest-HD group accumulated evidence more slowly (lower Drift rate) than controls in both the "few" condition (1.55 vs 2.37; probability that manifest-HD < healthy controls was more than 99%), which is relatively easy, and the more difficult "many[1]" condition (1.95 vs 2.91; probability that manifest-HD < healthy controls was more than 99%). Furthermore, the premanifest-HD group also demonstrated slower evidence accumulation than healthy controls, but only in the more difficult "many" condition (2.20 vs 2.91; probability that premanifest-HD < healthy controls was more than 99%), but not the "few" condition (2.04 vs 2.37; probability that premanifest-HD < healthy controls was 92%). Manifest and premanifest groups were similar in the "many" condition (1.95, 2.04, respectively; probability that manifest-HD < premanifest-HD was 84%). In contrast, for the "many" condition, the manifest-HD group was credibly slower than the premanifest-HD group (1.55 vs 2.04; probability that manifest-HD < premanifest-HD was 98%), who were not credibly different from controls (2.04 vs 2.37; probability that premanifest-HD < healthy controls is 92%). Within groups, only controls differed credibly in their evidence accumulation speed between the "few" and "many" conditions, with faster rates in the "many" condition (probability that "few" < "many" was 98%). The namesake parameter of the variability model, which provides a measure of the spread of individual scores within a group, was not found to be credibly different across groups; however, as there was still a substantial probability of differences. We assess the consequences of likely differences for clinical application in the supplement.

[1] The labels of "easy" and "difficult" are traditional in the experimental paradigms used with the RDM, but they do not exactly fit the task here. For instance, "few" trials that have a number of snowflakes close to the block's cutoff score would be quite difficult, in that many participants would be expected to give the incorrect label, while "many "trials with a large number of snowflakes would lead to a cognitively simple decision of choosing "many." However, the traditional labels of easy and difficult really aim to convey the cognitive load required of the participant to bare when encoding the stimuli on a trial, either by introducing more to process or by introducing more noise in the stimuli. In this sense, the "many" condition is "difficult" in that it has more snowflakes that need to be cognitively processed, while the "few" condition has less (and hence, "easier"). This difference in the average number of stimuli to process is really what defines the conditions, and this difference is captured in the higher drift rates (i.e., more to process) in the "many" condition.

Table 3.
*Group-level Mean (standard deviation) for Each RDM Parameter.*

| Group-Condition | Drift μ (mean) | Boundary α (mode) | Bias β | Non-Dec Time Rate η (mean) | Variability σ (mean) |
|---|---|---|---|---|---|
| MHD – Few | 1.55 (.15)[a,b] | 3.88 (.38) | .28 (.10) | 5.67 (1.37) [.176 s] | .67 (.13) |
| MHD – Many | 1.95 (.17)[c] | 3.94 (.39) | .25 (.04) | 5.67 (1.37) [.176 s] | .70 (.13) |
| PM – Few | 2.04 (.18)[b] | 3.54 (.30) | .24 (.03) | 7.74 (1.74) [.129 s] | .81 (.15) |
| PM – Many | 2.20 (.19)[d] | 3.51 (.31) | .25 (.03) | 7.74 (1.74) [.129 s] | .85 (.15) |
| HC – Few | **2.37 (.16)[a]** | 3.59 (.25) | .21 (.02) | 7.22 (1.55) [.138 s] | .83 (.13) |
| HC – Many | **2.91 (.19)[c,d]** | 3.89 (.27) | .19 (.02) | 7.22 (1.55) [.138 s] | .99 (.15) |

*Note.* Charting the point estimates of group-level parameters (MHD = manifest-HD Group; PM = premanifest-HD; HC = healthy controls). Measures of central tendency in parentheses indicate which was calculated from each parameter's posterior distribution, in accordance with the most appropriate measure given the shape of the posterior. The non-decision time was given an exponential distribution as a prior, which is parameterized by a rate; the mean of this distribution, and hence the mean non-decision time for each group, is 1/rate, and this mean is given in square brackets in seconds. Superscript letters correspond to pairwise credible differences as follows: (a) MHD was lower in drift than HC in Few; (b) MHD was lower than PM in Few; (c) HD was lower than HC in Many; (d) PM was lower than HC in Many. Bold indicates that HC showed a credible difference in drift between the Few and Many conditions.

### Correlations Between Model Parameters and Demographic and Clinical Variables

For correlation analyses, with the only two parameters that demonstrated acceptable reliability being evidence accumulation speed and response cautiousness, we combined the manifest- and premanifest-HD groups into a single group, given that these two groups are on a continuum for symptoms and demographic characteristics. We used Pearson product-moment correlations to examine associations between model parameters and key demographic, clinical, cognitive, and sleep quality variables. We did not apply a Bonferroni correction for multiple comparisons due to the exploratory nature of these analyses. For the combined HD group, neither evidence accumulation speed (Drift rates) nor response cautiousness (Boundary) parameters, for either the "few" or "many" conditions, were correlated with age, years of education, Disease Burden Score, or self-reported Total

Functional Capacity (all *ps* ≥ .064). In the healthy control group, however, higher response cautiousness in the "many" condition associated with older age ($p = .021$) and fewer years of education ($p = .037$), indicating that in "many" trials, older and less educated healthy control participants required more evidence before making a decision. No other correlations for the healthy control group were significant (all *p*s > .077).

To examine how model parameters related to performance on the HD-Mobile app tests, we examined correlations between evidence accumulation speed and response cautiousness, and outcome variables from the three HD-Mobile cognitive tasks (Table 4). For the HD group, we found that faster evidence accumulation speed in the "few" condition associated with faster Speeded Tapping ($p = .05$) and Two-choice Discrimination task performance ($p = .02$), as well as better on the Two-choice Discrimination ($p = .001$) and OILL tasks ($p = .001$). In effect, those HD participants with faster evidence accumulation on "few" trials were generally faster and more accurate on Two-choice Discrimination, and also faster in their responses on Speeded Tapping and more accurate on OILL. Faster evidence accumulation on "many" trials also associated significantly with greater overall percent correct on Two-choice Discrimination ($p = .002$). Also, the response cautiousness results indicated that HD participants who required more evidence (higher Boundaries) before a decision tended to be slower on Speeded Tapping ("few" and "many"trial *ps* = 0.011 and 0.015, respectively). In contrast, for the healthy control group, we observed few significant correlations. Specifically, those with faster evidence accumulation tended to be more accurate on the Two-choice Discrimination task ($p = .046$), but to have *slower* OILL response times ($p = .027$). A more cautious decision style for the "many" condition was associated with slower Two-choice Discrimination response times ($p = .026$).

For sleep and alertness measures, we found that for the HD group, participants with greater response cautiousness for both the "few" and "many" conditions also tended to report

themselves as experiencing more daytime sleepiness ($p$ = .009 and $p$ = .031, respectively). For the healthy control group, those with faster evidence accumulation in the "many" condition reported having better self-rated sleep quality ($p$ = .024) and higher alertness ($p$ = .008). Also, in control participants, greater response cautiousness in the "few" condition was associated with poorer self-reported sleep quality ($p$ = .015).

Table 4.
*Pearson Product-moment Correlations Between RDM Variables and HD-Mobile Cognitive Task Outcome Measures.*

| HDC ($n$ = 42) | Speeded Tapping RT | Two-choice % Correct | Two-choice RT | OILL % Correct | OILL RT |
|---|---|---|---|---|---|
| Drift Few | .31* | .52*** | .37* | .50*** | .15 |
| Drift Many | .23 | .49** | .21 | .28 | .12 |
| Boundary Few | -.39* | .16 | -.19 | .02 | .16 |
| Boundary Many | -.38* | .16 | -.26 | -.15 | .09 |
| HC ($n$ = 29) | | | | | |
| Drift Few | .06 | .21 | .13 | .21 | -.03 |
| Drift Many | .16 | .38* | -25 | .05 | .41* |
| Boundary Few | .02 | .11 | -.04 | .22 | -.13 |
| Boundary Many | .10 | .23 | -.42* | .05 | .34 |

*Note.* HDC = combined manifest- and premanifest HD group, HC = healthy control group, Speeded Tapping RT = 1/ln inter-tap interval in milliseconds on Speeded Tapping task, Two-choice % Correct = percent correct on the Two-choice Discrimination task, Two-choice RT = 1/ln of mean response time in milliseconds on Two-choice Discrimination task, OILL % Correct = overall percent correct on Object Identification and Location Learning task OILL RT = mean response time in milliseconds on the Object Identification and Location Learning task, *≤ .05, ** ≤ .01, *** ≤ .001. Because the relationship between RDM parameters and outcome measures (accuracy and mean response time) is highly non-linear, we do not expect perfect correlations between the two. The correlations between the RDM parameters and other task measures, which were not used to fit the model, provide some indication of the generalizability of the RDM parameters.

Table 5.

*Pearson Product-moment Correlations Between RDM Variables and Self-reported Sleep, Sleepiness, and Alertness Measures*

| HD-combined ($n$ = 42) | Sleep Disturbance | Sleep Quality | Daytime Sleepiness | Average Alertness |
|---|---|---|---|---|
| Drift Few | .02 | -.09 | -.10 | .14 |
| Drift Many | -.16 | .09 | -.18 | .05 |
| Boundary Few | .22 | -.16 | .40** | .14 |
| Boundary Many | .13 | -.09 | 33* | .08 |
| HC ($n$ = 29) | | | | |
| Drift Few | .10 | .15 | .07 | .06 |
| Drift Many | -.34 | -.42* | .23 | -.48** |
| Boundary Few | .26 | .43* | .11 | .15 |
| Boundary Many | -.23 | -.19 | .27 | -.33 |

*Note.* HD-combined = combined manifest- and premanifest HD group, HC = healthy control group, Sleep Disturbance = Sum of questions 1 – 5 from the SCOPA-SLEEP questionnaire (*lower* scored indicate better sleep), Sleep Quality = score from Q6 of SCOPA-SLEEP questionnaire (*lower* scores indicate better sleep), Daytime Sleepiness = sum of questions 7 – 12 from SCOPA-SLEEP Questionnaire (*lower* scores = less daytime sleepiness), Average Alertness = Average of responses on KSS questionnaire completed on Day one, four, and eight (*lower* scores = higher alertness), *≤ .05 ** ≤ .01.

## Discussion

In the first study to apply RDM analysis to two-choice decision making task data collected from smartphones, we detected group-level differences between premanifest- and manifest-HD and healthy control participants in cognitive parameters generated by the model, which were not apparent using observable performance, alone (McLaren et al., 2020). Because participants self-administered the task on smartphones, outside of the supervised clinic setting, this approach appears to suit the need for repeated longitudinal assessments to examine variability across time and in association to lifestyle factors such as sleep quality (Allard et al., 2014; Kaye et al., 2011; McCrae et al., 2012). Because we also observed differences in the premanifest-HD participant group, which overall differs only in small degrees from controls, an assessment method such as app-based self-assessment may also enable efficient and large-scale data collections, engendering the statistical power needed to observe the relatively small effects demonstrated by people in premanifest disease stages.

We expected, and observed, that the manifest-HD group would have slowed evidence accumulation (Drift Rate) in the RDM model due to their known slowing in speed of information processing (Stout et al., 2014) and because the evidence accumulation process is reliant on neural networks including the fronto-parietal pathway (Brosnan et al., 2020; Mulder, Van Maanen, & Forstmann, 2014) which are disrupted in manifest-HD (Poudel et al., 2014). This finding also adds speed of evidence accumulation to the cognitive processes showing detectable changes in premanifest HD (e.g., Beste et al., 2013; Labuschagne et al., 2013; Paulsen et al., 2013; Stout et al., 2014; You et al., 2014), which may have implications for the decision making in day-to-day functioning of those affected, especially situations in which information must be encoded and evaluated rapidly prior to making a decision, such as driving (Daneshi, Azarnoush, & Towhidkhah, 2020; Ratcliff, Smith, Brown, & McKoon, 2016). This observation highlights the importance of examining how evidence is accumulated and used in decision making in HD, which may also have implications for how people with HD respond to clinical interventions. The older age of the manifest-HD group did not account for the slower evidence accumulation speed; the premanifest-HD group showed the effect as well, yet they were of similar age to the control group. Further, previous research showed that older age is associated more with changes in decision cautiousness rather than evidence accumulation speed (Ratcliff, 2008). Moreover, the age regression models we tested demonstrated that evidence accumulation rates continued to differ between groups even when accounting for the effect of age (see the supplement for additional parameter analyses with age).

None of the group-level response cautiousness (Boundary) parameters were credibly different from one another, indicating that all groups required, on average, similar amounts of evidence before reaching a decision. We expected heterogeneity in response cautiousness values between groups because higher levels of striatal activation, the structure most heavily

affected in HD (Walker, 2007), is associated with reduced decision thresholds (Forstmann et al., 2008) and because impulsivity is observed in HD (Johnson et al., 2017; Mörkl et al., 2016). In particular, the decision cautiousness parameter is commonly associated with BOLD signals in the fronto-basal ganglia pathway (Mulder et al., 2014), which exhibits signs of atrophy in premanifest-HD (Stoffers et al., 2010). However, the exact direction of difference was difficult to predict due to the paucity of previous studies, and the unique pattern of neuronal pathology in HD, which limits the generalizability from other populations. We postulate the homogeneity in Boundary values may be secondary to our task instructions. Studies investigating Boundary values usually employ a speeded condition, instructing participants to respond as quickly as possible, and an accuracy condition, instructing accurate responses (e.g., Forstmann et al., 2008; Huang et al., 2015; Ratcliff, 2002), whereas we had only a speeded condition. Including speed *and* accuracy conditions may have enabled quantification of how, or whether, research participants adjust decision thresholds subsequent to direct instruction to decide rapidly or cautiously; indeed, there is strong evidence that speed/accuracy instructions affect the boundary in a selective fashion (Ratcliff, 2008). Instructing all participants to respond rapidly (i.e., somewhat impulsively) may have masked this symptom in the HD groups, if present, because controls may have also performed more quickly than they would have otherwise. As such, it will be informative in future studies to assess how and if decision cautiousness values change across conditions for premanifest- and manifest-HD and control groups when given both speed and accuracy instructions.

We did not find evidence of differences in response bias (Bias) or encoding/motor execution times (Non-decision Time) across conditions or groups, which we believe warrants further investigation given that HD is known to affect motor control negatively (Walker, 2007), which would have led to slower motor execution on the task. Alternatively, we may have been unable to observe effects in non-decision times because we were unable to come to

a satisfactory fit for this data due to convergence issues (see supplement for more details). Indeed, non-decision time was by far the most erratic parameter, in that its estimation was especially difficult and unreliable given the paucity of data (relative to other binary decision tasks; see the supplement for more details). This may be a natural consequence of clinical in-home assessment (Carpenter, Wycoff, & Trull, 2016). Perhaps this indicates that non-decision time is particularly difficult using remote ambulatory assessment, in that there are relatively more missing data points for some participants, which may limited available data for use in estimating the non-decision time parameter. Further model and protocol development in this direction would be helpful, especially as it may add to our understanding of the cognitive profile of HD.

We carefully audited the quality and robustness of our data and found data quality was sufficient for reliable group level comparisons for all variables, however for individual parameter values, we could use only evidence accumulation speed and response cautiousness data but not bias and encoding/motor execution time data for inferential statistics. The insufficient reliability of the Bias and Non-decision Time data was most likely due to some participants having large amounts of missing data (see the supplement for more on missing data). Our review of the raw data revealed that some participants appeared to rush through blocks of trials, repeatedly responding prior to the response window opening. Ensuring proper engagement and optimal effort is a known challenge for studies which have remote or unsupervised elements (Carpenter et al., 2016; Henson, Wisniewski, Hollis, Keshavan, & Torous, 2019) and so future iterations of the task should include methods to detect and prevent sub-optimal effort from participants.

We conducted exploratory correlational analyses between the two reliable individual estimate RDM variables and other key variables from our dataset to examine evidence for the validity and meaningfulness of the decision parameters generated from the modeling process.

The relationships between the evidence accumulation speed and response cautiousness variables and sleep-related variables were generally congruent with past literature. For example, disturbed sleep and more daytime sleepiness in our participants were associated with reductions in evidence accumulation rates and increases in response cautiousness, which has been observed in previous research (Patanaik, Zagorodnov, Kwoh, & Chee, 2014; Ratcliff & Van Dongen, 2009; Walsh, Gunzelmann, & Van Dongen, 2014). Interestingly, significant correlations generally emerged differently for the two groups. For example, for the HD group, self-reported daytime sleepiness, but not poorer self-reported night-time sleep quality, associated with increased response caution. Whereas for control participants poorer night-time sleep, but not daytime sleepiness, associated with response caution. Moreover, correlations between RDM variables and age and education were dissimilar for the HD and healthy control groups. These findings prompt the question of whether decision making processes of people with manifest- and premanifest-HD are affected differently by impaired sleep and fatigue, and by factors such as age and education, compared to the general population. In particular, understanding how the decision making of HD patients is affected by sleep disturbance and fatigue, which are common in HD (Herzog-Krzywoszanska & Krzywoszanski, 2019), should aid clinicians to guide management of cognitive symptoms of sleep disturbed and fatigued patients. The exploratory nature of these correlations only suggests this, however; future research should investigate further.

Future studies could implement longitudinal assessment on a time scale relevant to disease progression in HD (i.e., capturing declines in function across 12 to 24 months, Stout et al., 2012) to determine how well modeling of two-choice task performance in HD detects within-person disease progression. Inclusion of additional task conditions, e.g., speed and accuracy, and means of monitoring participant effort, will also address limitations noted in the present study. Another promising future direction would be to use the posteriors obtained

from the RDM data to predict group membership. For instance, one could have a person go through the same assessment, and while the model estimates this person's parameters, it could also estimate the probability of being in each group (control, premanifest-HD, manifest-HD; or in a clinical trial, on drug or placebo) through a signal detection method or otherwise a logistic regression that uses the posterior estimates from previous applications of the model.

In summary, our findings demonstrate applying RDM analysis to our two-choice decision making task data provides a highly sensitive method for detecting cognitive signs in both premanifest- and early manifest-HD. Without modeling the task data, we could not detect differences between the premanifest-HD and control groups (McLaren et al., 2020). We have shown, for the first time, that premanifest- and manifest-HD participants experience deficits in their evidence accumulation speed when processing visual stimuli during decision making. A final key aspect of the study is that data collection was conducted entirely outside of the laboratory, with participants self-administering the task on their own smartphones, which enables frequent and ongoing assessment without the limitation of geographical barriers. These combined factors mean this approach may be particularly useful for large longitudinal studies in HD and in post-marketing surveillance of ongoing use of new therapeutic drugs.

References

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*(1-2), 29-37.

Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., . . . Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PLoS ONE, 9*(12), e112197.

Bamford, K. A., Caine, E. D., Kido, D. K., Cox, C., & Shoulson, I. (1995). A prospective evaluation of cognitive decline in early Huntington's disease. *Functional and radiographic correlates, 45*(10), 1867-1873. doi:10.1212/wnl.45.10.1867

Beste, C., Stock, A. K., Ness, V., Hoffmann, R., Lukas, C., & Saft, C. (2013). A novel cognitive-neurophysiological state biomarker in premanifest Huntington's disease validated on longitudinal data. *Scientific reports, 3*. doi:10.1038/srep01797

Brosnan, M. B., Sabaroedin, K., Silk, T., Genc, S., Newman, D. P., Loughnane, G. M., . . . Bellgrove, M. A. (2020). Evidence accumulation during perceptual decisions in humans varies as a function of dorsal frontoparietal organization. *Nature human behaviour*, 1-12.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*: Sage.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253.

Campodonico, J. R., Aylward, E., Codori, A.-M., Young, C., Krafft, L., Magdalinski, M., . . . Brandt, J. (1998). When does Huntington's disease begin? *Journal of the International Neuropsychological Society, 4*(5), 467-473.

Carpenter, R. W., Wycoff, A. M., & Trull, T. J. (2016). Ambulatory assessment: New adventures in characterizing dynamic processes. *Assessment, 23*(4), 414-424.

Cosco, T. D., Doyle, F., Ward, M., & McGee, H. (2012). Latent structure of the Hospital Anxiety And Depression Scale: a 10-year systematic review. *Journal of Psychosomatic Research, 72*(3), 180-184.

Daneshi, A., Azarnoush, H., & Towhidkhah, F. (2020). A one-boundary drift-diffusion model for time to collision estimation in a simple driving task. *Journal of Cognitive Psychology, 32*(1), 67-81.

De Souza, J., Jones, L. A., & Rickards, H. (2010). Validation of self-report depression rating scales in Huntington's disease. *Movement Disorders, 25*(1), 91-96.

Eddy, C. M., & Rickards, H. E. (2013). Impact of cognitive and behavioural changes on quality of life in Huntington's disease. *Basal Ganglia, 3*(2), 123-126. doi:10.1016/j.baga.2013.01.085

Eddy, C. M., & Rickards, H. E. (2015). Cognitive deficits predict poorer functional capacity in huntington's disease: But what is being measured? *Neuropsychology, 29*(2), 268-273. doi:10.1037/neu0000134

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences, 105*(45), 17538-17542.

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*: Springer Science & Business Media.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science, 7*(4), 457-472.

Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., & Torous, J. (2019). Digital mental health apps and the therapeutic alliance: initial review. *BJPsych Open, 5*(1).

Herz, D. M., Little, S., Pedrosa, D. J., Tinkhauser, G., Cheeran, B., Foltynie, T., . . . Brown, P. (2018). Mechanisms Underlying Decision-Making as Revealed by Deep-Brain Stimulation in Patients with Parkinson's Disease. *Current Biology, 28*(8), 1169-1178.e1166. doi:https://doi.org/10.1016/j.cub.2018.02.057

Herzog-Krzywoszanska, R., & Krzywoszanski, L. (2019). Sleep disorders in Huntington's disease. *Frontiers in psychiatry, 10*, 221.

Huang, Y.-T., Georgiev, D., Foltynie, T., Limousin, P., Speekenbrink, M., & Jahanshahi, M. (2015). Different effects of dopaminergic medication on perceptual decision-making in Parkinson's disease as a function of task difficulty and speed–accuracy instructions. *Neuropsychologia, 75*, 577-587.

Johnson, P. L., Potts, G. F., Sanchez-Ramos, J., & Cimino, C. R. (2017). Self-reported impulsivity in Huntington's disease patients and relationship to executive dysfunction and reward responsiveness. *Journal of Clinical and Experimental Neuropsychology, 39*(7), 694-706. doi:10.1080/13803395.2016.1257702

Kaye, J. A., Maxwell, S. A., Mattek, N., Hayes, T. L., Dodge, H., Pavel, M., . . . Zitzelberger, T. A. (2011). Intelligent Systems for Assessing Aging Changes: Home-Based, Unobtrusive, and Continuous Assessment of Aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66B*(suppl 1), i180-i190. doi:10.1093/geronb/gbq095

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*: Academic Press.

Labuschagne, I., Jones, R., Callaghan, J., Whitehead, D., Dumas, E. M., Say, M. J., . . . Stout, J. C. (2013). Emotional face recognition deficits and medication effects in pre-manifest through stage-II Huntington's disease. *Psychiatry Research, 207*(1-2), 118-126. doi:10.1016/j.psychres.2012.09.022

Marinus, J., Visser, M., van Hilten, J. J., Lammers, G. J., & Stiggelbout, A. M. (2003). Assessment of Sleep and Sleepiness in Parkinson Disease. *Sleep, 26*(8), 1049-1054. doi:10.1093/sleep/26.8.1049

Maroof, D. A., Gross, A. L., & Brandt, J. (2011). Modeling longitudinal change in motor and cognitive processing speed in presymptomatic Huntington's disease. *Journal of Clinical and Experimental Neuropsychology, 33*(8), 901-909. doi:10.1080/13803395.2011.574606

McCrae, C. S., Vatthauer, K. E., Dzierzewski, J. M., & Marsiske, M. (2012). Habitual sleep, reasoning, and processing speed in older adults with sleep complaints. *Cognitive therapy and research, 36*(2), 156-164.

McLaren, B., Andrews, S. C., Glikmann-Johnston, Y., Mercieca, E.-C., Murray, N. W. G., Loy, C., . . . Stout, J. C. (2020). Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *Journal of Neurology*.

Mörkl, S., Müller, N. J., Blesl, C., Wilkinson, L., Tmava, A., Wurm, W., . . . Painold, A. (2016). Problem solving, impulse control and planning in patients with early- and late-stage Huntington's disease. *European Archives of Psychiatry and Clinical Neuroscience, 266*(7), 663-671. doi:10.1007/s00406-016-0707-4

Mulder, M., Van Maanen, L., & Forstmann, B. (2014). Perceptual decision neurosciences–a model-based review. *Neuroscience, 277*, 872-884.

Patanaik, A., Zagorodnov, V., Kwoh, C. K., & Chee, M. W. (2014). Predicting vulnerability to sleep deprivation using diffusion model parameters. *Journal of Sleep Research, 23*(5), 576-584.

Paulsen, J. S., Smith, M. M., & Long, J. D. (2013). Cognitive decline in prodromal Huntington disease: Implications for clinical trials. *Journal of Neurology, Neurosurgery and Psychiatry, 84*(11), 1233-1239. doi:10.1136/jnnp-2013-305114

Penney, J. B., Vonsattel, J. P., Macdonald, M. E., Gusella, J. F., & Myers, R. H. (1997). CAG repeat number governs the development rate of pathology in Huntington's disease. *Annals of Neurology, 41*(5), 689-692.

Poudel, G. R., Stout, J. C., Domínguez D, J. F., Salmon, L., Churchyard, A., Chua, P., . . . Egan, G. F. (2014). White matter connectivity reflects clinical and cognitive status in Huntington's disease. *Neurobiology of Disease, 65*, 180-187. doi:https://doi.org/10.1016/j.nbd.2014.01.013

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic bulletin & review, 9*(2), 278-291.

Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging, 23*(4), 900.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences, 20*(4), 260-281.

Ratcliff, R., & Van Dongen, H. P. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic bulletin & review, 16*(4), 742-751.

Ross, C. A., Aylward, E. H., Wild, E. J., Langbehn, D. R., Long, J. D., Warner, J. H., . . . Tabrizi, S. J. (2014). Huntington disease: Natural history, biomarkers and prospects

for therapeutics. *Nature Reviews Neurology, 10*(4), 204-216. doi:10.1038/nrneurol.2014.24

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 485-493.

Starkstein, S. E., Brandt, J., Bylsma, F., Peyser, C., Folstein, M., & Folstein, S. E. (1992). Neuropsychological correlates of brain atrophy in Huntington's disease: a magnetic resonance imaging study. *Neuroradiology, 34*(6), 487-489. doi:10.1007/BF00598956

Stoffers, D., Sheldon, S., Kuperman, J. M., Goldstein, J., Corey-Bloom, J., & Aron, A. R. (2010). Contrasting gray and white matter changes in preclinical Huntington disease: an MRI study. *Neurology, 74*(15), 1208-1216. doi:10.1212/WNL.0b013e3181d8c20a

Stout, J. C., Jones, R., Labuschagne, I., O'Regan, A. M., Say, M. J., Dumas, E. M., . . . Frost, C. (2012). Evaluation of longitudinal 12 and 24 month cognitive outcomes in premanifest and early Huntington's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 83*(7), 687-694. doi:10.1136/jnnp-2011-301940

Stout, J. C., Queller, S., Baker, K. N., Cowlishaw, S., Sampaio, C., Fitzer-Attas, C., . . . Thompson, J. (2014). HD-CAB: A cognitive assessment battery for clinical trials in Huntington's disease. *Movement Disorders, 29*(10), 1281-1288. doi:10.1002/mds.25964

Tabrizi, S. J., Scahill, R. I., Owen, G., Durr, A., Leavitt, B. R., Roos, R. A., . . . Langbehn, D. R. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: Analysis of 36-month observational data. *The Lancet Neurology, 12*(7), 637-649. doi:10.1016/S1474-4422(13)70088-7

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*(1), 44.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Epskamp, S. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review, 25*(1), 35-57.

Walker, F. O. (2007). Huntington's disease. *The Lancet, 369*(9557), 218-228.

Walsh, M., Gunzelmann, G., & Van Dongen, H. (2014). *Comparing accounts of psychomotor vigilance impairment due to sleep loss.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology, 54*(1), 39-52.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in neuroinformatics, 7*, 14.

You, S. C., Geschwind, M. D., Sha, S. J., Apple, A., Satris, G., Wood, K. A., . . . Possin, K. L. (2014). Executive functions in premanifest Huntington's disease. *Movement Disorders, 29*(3), 405-409. doi:10.1002/mds.25762

Zhang, Y., Long, J. D., Mills, J. A., Warner, J. H., Lu, W., Paulsen, J. S., . . . Coordinators of the Huntington Study, G. (2011). Indexing disease progression at study entry with individuals at-risk for Huntington disease. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics, 156B*(7), 751-763. doi:10.1002/ajmg.b.31232

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

# Chapter Five: Study Four

# Greater Time in Bed and Less Physical Activity Associate with Poorer Cognitive Functioning Performance in Huntington's Disease

## 5.1. Explanatory Note

The research presented in this chapter addresses my final research aim, which was adding to the limited literature on how sleep and physical activity habits relate to cognitive functioning in HD. The study presented in Chapter Two gave the first evidence that remote assessment of cognition, sleep, and physical activity in Huntington's disease was feasible. Then, the studies presented in Chapters Three and Four showed that the HD-Mobile cognitive tasks generated cognitive data that was reliable, valid, and scientifically and clinically meaningful. Study Four represented the culmination of the overall research project because it combined all aspects of our remote assessment protocol and provided insight into how sleep, physical activity, and cognition associate in people with the expanded HD repeat as they live in their homes and communities. Study Four extended the findings of previous research in HD by assessing sleep, physical activity, and cognition in the one study and by adding ecological relevance to the findings because some data were collected in real time, rather than using retrospective self-report, and all assessments were conducted in the home.

**Note.** This chapter was submitted to the journal *SLEEP* and is presented in its submitted format.

Greater time in bed and less physical activity associate with poorer cognitive functioning performance in Huntington's disease

Brendan McLaren[1], BBioSc (Hons), GradDipPsych, BSc (Hons), ORCID iD 0000-0002-8769-8629, Sean P. A. Drummond, BSc, PhD, ORCID iD 0000-0002-9815-626X ,Yifat Glikmann-Johnston[1], BA, MA, PhD, ORCID iD 0000-0002-6128-7768, Clement Loy[2,3], BA, MBBS, MMed (Clin Epi), FRACP, PhD, Mark A. Bellgrove[1], BSc (Hons), PhD ORCID iD 0000-0003-0186-8349, Julie C. Stout[1], BSc, MA, PhD[1], ORCID iD 0000-0001-9637-8590, Sophie C. Andrews[1,4,5], BA, GradDipPsych, BSc (Hons), DPsych (Clin Neuro), ORCID iD 0000-0002-2103-7748,

[1]*Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, Victoria, Australia*

[2]*Brain and Mind Centre, the University of Sydney, Sydney Australia*

[3]*Huntington Disease Service, Westmead Hospital, Sydney Australia*

[4]*Neuroscience Research Australia, Sydney, NSW, Australia*

[5]*School of Psychology, University of New South Wales, Sydney, Australia*

**Correspondence to**: Dr Sophie Andrews, NeuRA, Margarete Ainsworth Building, Barker Street, Randwick, NSW 2031, Australia; Email: s.andrews@neura.edu.au; Phone: +61 2 9399 1076

## Abstract

### Study Objectives

This study aimed to investigate how sleep and physical activity habits related to cognitive functioning, in naturalistic settings, in early Huntington's disease (HD).

### Methods

Forty-two participants with the expanded HD repeat (20 manifest, 22 premanifest) and 29 healthy controls wore Fitbit One sleep and activity monitors for seven days and seven nights. They used a smartphone application to complete daily sleep and activity diaries, sleep and mood inventories, and a brief battery of cognitive tests, which were completed on day 8 of the study. All data was collected in naturalistic home and community settings.

### Results

Amongst participants with the expanded HD gene, greater time spent in bed, measured by Fitbit, was associated with poorer accuracy and response speed on a test of visual memory, whereas lower levels of physical activity, measured by Fitbit, was associated with poorer accuracy on a test of executive functioning and working memory. Neither time in bed nor physical activity associated with a test of psychomotor speed. Groups were mostly similar across a range of Fitbit and self-report measures of sleep and physical activity, although the Manifest-HD group spent more time in bed than the Premanifest-HD and Healthy Control groups, and self-reported better sleep quality and more time spent sitting than the Healthy Control group and the Premanifest-HD group, respectively.

### Conclusions

Sleep timing and physical activity relate to cognitive functioning in HD and may be important targets for management in behavioural intervention studies aimed at improving cognition in HD.

Cognitive impairments are a primary feature of Huntington's disease HD; [1], and limit day to day functioning and quality of life [2,3]. Disordered sleep and sub-optimal levels of physical activity have been consistently associated with poorer cognitive functioning across the lifespan in the general population [4-11], and are known risk factors for the development of dementias including Alzheimer's and Parkinson's disease dementias[12-14]. In comparison, although previous studies have separately investigated interactions between sleep and cognition, and physical activity and cognition [15-17] in HD, the relationship between sleep, physical activity, and their relative association with cognition in HD has not been examined. Measuring both sleep and physical activity when assessing their relationship to cognitive functioning is an important step because the three factors have complex interactions[18-20]. For example, exercise can benefit cognition directly, but also indirectly by improving sleep quality which can then have flow on benefits for cognitive functioning[18,20]. Past studies that assessed sleep-cognition and physical activity-cognition interactions in HD measured cognitive performance in the lab, rather than naturalistic settings, limiting the generalizability of findings to real world functioning. In the current study, we aimed to address these gaps by providing a more comprehensive picture of how sleep and physical activity jointly relate to cognitive functioning and by measuring all factors in naturalistic home and community settings. Understanding these relationships will take an important step toward empirically based lifestyle interventions and better symptom management for patients.

Sleep disturbance in HD is common and known to be related to cognition [21-23]. Estimated rates of sleep disturbance in HD vary depending on assessment method and metric of measurement, however, it has been reported that close to 90% of HD patients will report experiencing sleep problems [24]. Consistently higher rates of sleep disturbance are present in premanifest- and manifest- HD patients than in healthy controls or non-affected family members, and in people with manifest-HD compared to premanifest-HD, indicating

progressively worsening sleep with advancing disease [15,21,22,25,26]. Common early sleep symptoms include changes in circadian rhythm, increased awakenings at night, and increased time to first rapid eye movement (REM) episode [21,27,28]. The limited research available on relationships between sleep and cognition suggests that a later habitual wake time correlates with poorer cognitive functioning [15,16]. Specifically, Aziz and colleagues[15] (*n* = 21 participants with premanifest-HD and *n* = 37 participants with manifest-HD ) and Diago and colleagues[16] (with *n* = 23 participants with premanifest-HD and *n* = 15 with manifest-HD) reported that those who woke later in the morning had worse cognitive functioning across tests of processing speed and higher level attention and inhibition. Combining the results of both studies shows that habitual wake time relates to cognition for both people with premanifest- and manifest-HD. Nonetheless, more work is needed to fully understand the relationships between sleep timing and habits and cognition in HD.

In terms of physical activity, only one previous observational study examined relationships between physical activity and cognition in HD [17]. Physical activity was measured with Fitbit Ultra activity monitors and calculated by combining mean steps, distance walked, and stairs climbed over a three day period. The participants, 48 people with the expanded HD repeat (44 in the premanifest stage and four in the early manifest stage), who engaged in more physical activity, performed better on the Symbol Digit Modalities Test (SDMT), a measure of processing speed sensitive to cognitive symptom progression in HD [29]. In contrast, level of physical activity was not associated with performance on the Stroop Colour and Word Test, a measure of higher-level attention. This suggests physical activity levels may not associate equally with all domains of cognition, which is consistent with findings in non-clinical samples [4,30,31]. Further, a meta-analysis of physical activity interventions in HD identified six studies that utilized cognitive measures to assess intervention outcomes; although the results were mixed, few significant improvements in

cognition were found [32]. Given these mixed results, there is a clear need to investigate the relationships between physical activity and cognition in HD across both the premanifest and manifest stages and with a broader range of cognitive tests.

In this study, we examined how sleep and physical activity relate to the day-to-day cognitive functioning in premanifest- and manifest-HD. For the assessment of sleep and physical activity, we used both passive activity monitoring and subjective (self-report) measures, and related these to cognitive measures from a self-administered mobile application [33] in people within their everyday settings of home and their communities. Previously, only a very small set of cognitive tasks have been used when assessing sleep-cognition and physical activity-cognition interactions in HD. These have mainly consisted of the three tests included in the Unified Huntington's Disease Rating Scale (UHDRS [34]) cognitive battery (i.e., Symbol Digit Modalities Test, Phonemic Fluency, and Stroop Color-Word Interference). The current study extends this by utilising new cognitive tasks and broadening the range of domains assessed in relation to sleep and physical activity habits in HD. We also used a combination of passive and subjective measures of because active and self-report sleep measures relate to separate sleep parameters[35]. We had two primary aims. The first was to understand how sleep and physical activity habits, assessed jointly using activity recording, in real time, rather than separate studies, in natural rather than clinical settings, relates to cognitive functioning in people with the expanded HD repeat. The second aim was to widen the range of measures used simultaneously to characterise sleep and physical activity in HD, and to investigate an alternative data collection method that enables self-assessment of cognition in proximity to actual sleep and physical activity by employing a mobile app.

## Method

### Participants

We studied forty-two people with the expanded HD repeat (20 manifest, 22 premanifest), and 29 healthy control participants (Table 1) across eight days. For inclusion in the study, we required participants to be 18 – 70 years of age, fluent in English, and to own an Android or Apple smartphone running at least Android 4.0 or iOS 7 operating systems. We excluded participants who were currently enrolled in any clinical drug trial, had a history of neurological illness (other than HD), psychiatric disorder other than depression or anxiety (both are common features of HD [1]), significant head injury, or substance abuse. To be included in the HD CAG expansion group, participants had to have CAG expansion of 39 or more. Further, to be in the Manifest-HD group, participants were required to have been diagnosed with HD. To be included in the Premanifest-HD group, we required participants to have disease burden scores (DBS) of ≥ 200 calculated as age x [CAG – 35.5] [36]. For additional clinical characterisation, we also calculated the CAG-Age product (CAP score) for Premanifest and Manifest HD participants as age x [CAG – 33.6] [37]. The Manifest-HD group was older than the Premanifest-HD and Healthy Control groups ($p$ = .005 and .027, respectively), and had lower self-reported total functional capacity (TFC) [38] than Premanifest-HD and Healthy Control groups (both $ps$ < .001), as expected. The Manifest-HD group had higher UHDRS total motor scores (TMS), DBS and CAG-Age product scores than the Premanifest-HD group (all $ps$ < .008). The Premanifest-HD group had fewer years of formal education than the Healthy Control group ($p$ = .012). The groups did not differ in their anxiety or depression scores and the Premanifest- and Manifest-HD groups did not differ in average CAG expansion length. We recruited participants through participant databases at Monash University, Melbourne, and Westmead Hospital, Sydney, which included volunteers who had consented to being contacted about research participation opportunities.

Table 1.

*Group-level Comparisons for Participant Characteristics and Subjective and Objective Sleep and Physical Activity Variables.*

| Characteristic | Manifest-HD | Premanifest-HD | Healthy Controls | Group differences (*p*-value) and effect size, Cohens *d* | | |
|---|---|---|---|---|---|---|
| *n* | 20 | 22 | 29 | MHD - PHD | MHD – HC | PHD - HC |
| Age (Mean Years (SD), Range) | 51.60 (8.03)<br>31 - 65 | 42.09 (8.52)<br>21 - 58 | 44.31 (10.93)<br>28 - 65 | ***p* = .005**<br>*d* = 1.15 | ***p* = .027**<br>*d* = 0.76 | *P* = .69<br>*d* = 0.23 |
| Education (Mean Years (SD), Range) | 14.33 (2.14)<br>12 - 18 | 13.25 (2.09)<br>10 - 18 | 14.88 (1.68)<br>12 -19 | *p* = .18<br>d = *0.51* | *p* = .59<br>d = 0.29 | *p* = .012<br>*d* = 0.86 |
| CAG Repeat length (Mean Years (DS), Range) | 42.88 (2.11)<br>41 - 49 | 42.95 (2.16)<br>41 - 51 | - | *p* = .89<br>*d* = 0.03 | - | - |
| Disease-burden score (Mean (SD), Range) | 379.64 (58.85)<br>275.00 – 484.50 | 304.12 (56.13)<br>217.50 – 435.00 | - | ***p* < .001**<br>*d* = 1.31 | - | - |
| CAG-Age Product (Mean (SD), Range) | 475.62 (60.78)<br>367.00 – 589.38 | 382.47 (66.45)<br>270.86 – 541.72 | - | ***p* < .001**<br>*d* = 1.46 | - | - |
| HADS Anxiety (Mean (SD), Range) | 5.85 (3.62)<br>2 – 17 | 6.36 (4.29)<br>0 – 15 | 6.00 (4.53)<br>0 – 16 | *p* = .92<br>*d* = 0.13 | *p* = .99<br>*d* = .04 | *p* = .95<br>*d* = .08 |
| HADS Depression (Mean (SD), Range) | 5.40 (4.45)<br>0 – 14 | 3.50 (3.26)<br>0 – 11 | 3.17 (3.42)<br>0 – 16 | *p* = .23<br>*d* = 0.49 | *p* = .10<br>*d* = 0.56 | *p* = .95<br>*d* = .10 |
| UHDRS TFC (Self-report) (Mean (SD), Range) | 9.26 (2.42)<br>6 - 13 | 12.64 (0.90)<br>9 - 13 | 12.97 (0.19)<br>12 - 13 | ***p* < .001**<br>*d* = 1.86 | ***p* < .001**<br>*d* = 2.16 | *p* = .69<br>*d* = 0.51 |
| UHDRS TMS* (Self-report) (Mean (SD), Range) | 15.89 (12.59)<br>1 – 40<br>(*n* = 7) | 1.25 (2.55)<br>0 – 7<br>(*n* = 12) | - | ***p* = .006** | - | - |

*Note.*, HADS = Hospital Anxiety and Depression Scale, UHDRS TFC = Unified Huntington's Disease Rating Scale Total Functional Capacity, UHDRS TMS = Unified Huntington's Disease Rating Scale Total Motor Score, which was only available for a subset of participants.

We also recruited people with the expanded HD repeat and healthy control participants using advertisements for the study posted on Facebook and at local HD community events. (See Table 1 for participant characterisation).

**Equipment and Measures**

We used a smartphone application, HD-Mobile [39], to measure cognition and for participants to report on their sleep, physical activity, and mood. HD-Mobile was designed for touch screen smartphones running Android 4.0 and iOS 7 (and above) operating systems. The app includes three cognitive tasks and a questionnaire feature that was used for self-report measures. In total, we collected data on three self-report questionnaires and three cognitive tasks using HD-Mobile, which included the Object Information and Location Learning (referred to as the Visual Memory task in this paper), Two-choice Discrimination, and Speeded Tapping. We have previously described the tasks in detail [39]. Briefly, the *Visual Memory task* assesses learning and memory for visual stimuli in two conditions, object identity and object location. For this study, we recorded the overall accuracy (percent correct) and mean response speed (in milliseconds) and used both in subsequent analyses. The *Two-choice Discrimination task* assesses perceptual decision making and has a working memory component as participants are required to learn from feedback accumulated across trials to improve their accuracy. For this task, participants view a scene depicting snowflakes falling from the sky, and are instructed to quickly decide, without counting, whether there are "few" or "many" snowflakes on the screen. Participants receive feedback, in the form of a "happy face" for correct decisions and a "sad face" for incorrect decisions, which they can use to adjust their subsequent decisions. Overall accuracy (percent correct) and response speed (milliseconds) were the dependent variables. *Speeded Tapping* is a bimanual speeded finger-tapping task that requires the participant to tap buttons on the smartphone screen as fast as

possible with alternating thumbs, providing a measure of psychomotor speed. The dependent variable was mean inter-tap interval (milliseconds), termed "tapping speed".

The subjective self-report questionnaires included on HD-Mobile included the SCale of Outcomes in Parkinson's disease – Sleep (SCOPA –SLEEP) [40], the Hospital Anxiety and Depression Scale HADS [41], and a sleep and activity diary that we designed for the study. The SCOPA-SLEEP scale was originally developed for the measurement of night-time sleep (5 items) and daytime sleepiness (6 items) in the past month in patients with Parkinson's disease, and has shown good reliability, $\alpha = .89$ for night-time sleep items and .85 for daytime sleepiness items, and validity when used with HD patients [15]. Higher scores for each of the subscales indicate poorer night-time sleep and higher levels of daytime sleepiness. The HADS is a widely used and reliable [41-43] 14-item self-report questionnaire, which assesses feelings of anxiety and depression over the previous week. The sleep and activity diary included a standard sleep diary and questions about physical activity. The sleep diary required participants to report the time they closed their eyes with the intention of sleeping on the previous night, the time they woke that morning, and to rate how well they slept on the previous night, using a seven-point Likert scale (very well to very badly; lower scores indicating *better* sleep). For the activity diary, participants reported the amount of time they spent engaging in moderate to vigorous physical activity on the previous day and also estimated the time they spent sitting. We provided participants with examples of moderate and vigorous physical activity to use as a guide to anchor their ratings.

We used the Fitbit One to assess sleep and physical activity objectively. The Fitbit One is a lightweight, portable, activity monitor that utilizes a triaxial accelerometer to measure total step count, total activity calories, minutes spent in sedentary behaviours during the day, as well as aspects of sleep such as time in bed, total sleep time, and sleep efficiency [44,45]. The Fitbit One has good validity for detecting total steps and energy expenditure, and is

reliable as a measure of sleep habits (time in bed and wake/sleep schedules) in the general population [17,44,46] and in HD, although it is not recommended as a measure of total sleep time or sleep efficiency in HD due to unsatisfactory agreement with benchmark polysomnography and EEG measures [47,48]. From the Fitbit we used "time in bed" and "activity calories" as measures of sleep and physical activity habits in our analyses because they are reliable measures in HD and because both sleep habits and activity levels have previously been linked to cognitive functioning in HD [15-17]. For both time in bed and activity calories, we calculated the mean for each participant over the seven days and nights of the study. For inclusion in the analysis, each participant had to have at least three days, or nights, of available data.

To measure sleep characteristics using the Fitbit One, participants set the device to "sleep mode" when they were ready to go to sleep, and then manually ended sleep mode once they had awoken and planned to get up for the day, thereby providing a measure of the amount of time the person spent attempting to sleep. We used this variable, referred to as "time in bed", in subsequent analyses. In some cases, the FitBit data indicated that a participant had neglected to end sleep mode on arising from bed (i.e., clear activity periods commenced in the morning following the inactivity of sleep or clear activity at night after setting sleep mode). In these cases, we manually adjusted the end period of sleep based on the combination of FitBit activity data and the participant's sleep diary. We estimate this occurred in approximately 30% of entries for participants with the HD gene expansion and 23% of entries for participants in the Healthy Control group. Note that because in some cases we used the self-report data to refine the Fitbit sleep estimates, these two methods were non-independent, and it was therefore not possible to directly study their agreement using correlations. This procedure is similar to methods used in contemporary sleep studies [49-51] and helps to ensure the accuracy of the time in bed measure by removing the effect of operator error, such as participants forgetting to turn sleep mode off after waking. In addition to time

in bed, we also obtained a measure of habitual sleep and wake times from the Fitbit. We calculated this as the average sleep and wake time taken over the seven nights and mornings of the study. We required participants to have at least three days of usable data available before calculating and using their sleep or physical activity data in any analyses. Overall, four participants from the healthy control group and two participants with the expanded HD repeat did not have sufficient Fitbit time in bed or activity calorie data.

**Procedure**

We provided participants with both video and written instructions to enable them to install HD-Mobile on their smartphones and, and to operate the Fitbit devices. The instructions described what participants were required to do each day of the study. Participation occurred across an eight-day period, during which participants were required to complete one or more study procedures each day. We instructed participants to complete tasks and questionnaires within an hour of waking on each day of the study. This was to ensure the sleep diary was completed soon upon awakening, while the memory of their previous night's sleep was fresh, and to try to ensure all participants completed cognitive tasks at a similar point in their circadian cycles. On the morning of Day 1, participants logged into HD-Mobile and completed an electronic informed consent form, the three cognitive tasks, all of the questionnaires, and attached the Fitbit One. The sleep and activity diary was completed on all days of the study and the cognitive tasks were completed again on Days 4 and 8. For this study we only analysed the cognitive data collected on Day 8 because we were interested in how seven days of sleep and physical activity would relate to cognition on the eighth day. Participants continued wearing the Fitbit until the morning of Day 8. Total time commitment for the active parts of the study was two and a half hours, although we collected data passively using the Fitbit for seven days and nights.

**Data Analysis**

We compared Manifest-HD, Premanifest-HD, and Healthy Control groups on demographic, cognitive, and mood variables and both subjective and passively derived sleep and physical activity variables using a series of one-way analyses of variance (ANOVA) both subjective and passive monitoring of sleep and PA with Tukey's Honestly Significant Difference post-hoc tests. Note that we did not study the concordance between the self-report and Fitbit measures of sleep and physical activity because these measures differed in terms of the specific constructs they were intended to measure, or in the case mentioned above of sleep timing, the measures were not independent. We have previously presented group level comparisons of HD-Mobile cognitive task data which included analysis of the cognitive data which was collected on Days 1, 4, and 8 [33]. The cognitive data presented below are from Day 8 only, as explained above.

We computed five separate hierarchical multiple regression models to determine the effects of sleep and physical activity (tested as independent variables within the model) on performance on the five cognitive outcome measures from the HD-Mobile in the participants with the expanded HD only. In addition to the sleep and physical activity independent variables, each model also included DBS and the HADS depression subscale score (HADS-D) as the first step, to account for effects of depression symptoms and estimated disease progression on cognition, which was followed by entering mean activity calories and time in bed in the second step. We initially included a time in bed by activity calories interaction term to assess if the sleep and physical activity measures interacted in their associations with cognitive performance. The interaction term was non-significant in all models and generally weakened each model. As such, the regression models below are presented without the interaction term. Statistical analyses were performed using IBM SPSS V. 26 software.

Missing data were excluded pairwise from the regression analyses. Alpha was set at .05 for all analyses.

## Results

Our main goal was to examine the joint relationships of sleep and physical activity levels in relation to cognition in patients with the expanded HD repeat. To provide a context for these results, we first describe group comparisons on subjective and passive sleep and physical activity measures, and performance on the cognitive tasks. We then turn to our main analyses, where we model the relative associations of sleep and physical activity with mobile app based cognitive measures.

### Group Level Comparisons of Sleep, Physical Activity, and Cognitive Performance

We report comparisons for passive measures of sleep and physical activity first. In terms of the Fitbit sleep findings, the Manifest-HD group spent on average more time in bed than both the Premanifest-HD and Healthy Control groups. Habitual sleep times were very similar across groups, averaging just before 11 pm; average wake times were more variable, with averages across the group ranging from just before 7 am in the Healthy Controls and closer to 7:40 am in the Manifest-HD group, with no significant differences, but the Manifest-HD group showed a statistical trend ($p$ =.056) for arising later than the Healthy Controls. See Table 2 for all sleep and physical activity variable comparisons. There were no significant differences in Fitbit physical activity measures between the three groups. The groups were similar across two self-report sleep outcomes from the SCOPA-SLEEP questionnaire, night time sleep disturbance and daytime sleepiness. In contrast, the Manifest-HD group's scores on the sleep diary rating of sleep somewhat contradicted their SCOPA-SLEEP scores. They self-reported *better* sleep than the Healthy Control group on the diary measure, but not on SCOPA-SLEEEP questions. In terms of physical activity, diary-based records of minutes of moderate to vigorous physical activity were highly variable but did not

differ significantly between groups. In contrast, diary-based records indicated more sitting time reported by Manifest-HD compared to both Premanifest-HD and Healthy Control groups. This contrasted somewhat with Fitbit data which indicated time spent sedentary was similar for all three groups.

In terms of cognitive data collected on Day 8, as expected the Manifest-HD group performed significantly worse than Healthy Controls on all cognitive outcomes except on the accuracy measure of the Visual Memory task where they performed worse than the Premanifest-HD group. The Healthy Control group were intermediate on Visual Memory and not significantly different from either other group. The Manifest HD group was also slower than the Premanifest-HD group on Speeded Tapping. There were no significant differences between Premanifest-HD and Healthy Control participants on any of the cognitive measures. A summary of group-level cognitive performances on the HD-Mobile cognitive tasks, as measured on Day 8 of the study, is presented in Table 2.

Table 2.

*Group-level Comparisons for Passive and Subjective Sleep and Physical Activity Variables and HD-Mobile Cognitive Variables.*

| Characteristic | Manifest-HD | Premanifest-HD | Healthy Controls | Group differences (*p*-value) and effect size, Cohens *d* | | |
|---|---|---|---|---|---|---|
| Sleep variables | | | | | | |
| SCOPA-NS (Mean (SD), Range) | 4.90 (3.28) 0 - 15 | 4.68 (2.59) 0 - 9 | 5.28 (2.80) 0 - 11 | *p* = .97 *d* = 0.07 | *p* = .90 *d* = 0.12 | *p* = .75 *d* = 0.22 |
| SCOPA-DS (Mean (SD), Range) | 2.40 (2.96) 0 – 9 | 2.14 (1.89) 0 - 8 | 1.66 (1.01) 0 - 3 | *p* = .90 *d* = 0.18 | *p* = .41 *d* = 0.42 | *p* = .67 *d* = 0.32 |
| Sleep diary rating of sleep (Mean (SD), Range) | 2.45 (0.85) 1.0 – 4.50 | 2.79 (0.78) 1.00 – 4.13 | 3.04 (0.81) 1.63 - 4.63 | *p* = .37 *d* = 0.41 | ***p* = .038** *d* = 0.71 | *p* = .52 *d* = 0.32 |
| Fitbit time in bed (Mean (SD), Range) | 9:03.29 (0:58.17) 07:36.25 – 11:00.42 | 8:09.14 (52.05) 05:57.00 – 09:34.08 | 7:53.24 (52.19) 05:32.40 – 09:31.17 | ***p* = .007** *d* = 0.98 | ***p* < .001** *d* = 1.27 | *p* = .60 *d* = 0.30 |
| Fitbit habitual sleep time (Mean (SD), Range) | 22:55.20 (0:51.38) 21:44.17 – 00:25.34 | 22:51.04 (0:58.10) 21:07.17 – 00:11.34 | 22:59.17 (1:09.24) 20:49.42 – 02:00:10.20 | *p* = .98 *d* = 0.08 | *p* = .98 *d* = 0.06 | *p* = .91 *d* = 0.13 |
| Fitbit habitual wake time (Mean (SD), Range) | 7:36.50 (1:09.06) 05:37.42 – 10:04.34 | 7:05.29 (0:40.37) 05:50.42 – 07:55.51 | 6:51.29 (0:59.28) 05:19.08 – 09:00.59 | *p* = .28 *d* = 0.55 | *p* = .056 *d* = 0.71 | *p* = .71 *d* = 0.28 |
| Physical activity variables | | | | | | |
| Activity diary: Minutes of moderate and vigorous activity Mean (SD), Range | 74.85 (66.39) 0.00 -255.00 | 130.39 (144.04) 3.75 – 487.50 | 84.22 (108.28) 0.00 – 412.50 | *p* = .25 *d* = 0.50 | *p* = .96 *d* = 0.10 | *p* = .32 *d* = 0.36 |
| Activity diary: Minutes sitting Mean (SD), Range | 513.21 (182.17) 240.00 – 832.50 | 377.31 (160.00) 157.50 – 750.00 | 356.48 (161.86) 125.63 – 690.00 | ***p* = .028** *d* = 0.79 | ***p* = .005** *d* = 0.91 | *p* = .90 *d* = 0.13 |
| Fitbit minutes sedentary (Mean (SD), Range) | 619.44 (110.31) 450.71 – 807.43 | 615.74 (134.00) 254.43 – 886.00 | 664.17 (92.74) 507.00 – 798.00 | *p* = .99 *d* = 0.03 | *p* = .38 *d* = 0.44 | *p* = .33 *d* = 0.42 |
| Fitbit activity calories Mean (SD), Range | 1224.00 (501.02) 333.33 – 1936.43 | 1329.42 (473.97) 741.14 – 2452.00 | 1120.07 (354.07) 588.57 – 1814.57 | *p* = .73 *d* = 0.22 | *p* = .71 *d* = 0.24 | *p* = .26 *d* = 0.50 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Fitbit steps<br>Mean (SD),<br>Range | 8413.26 (3819.36)<br>1799.00 – 13838.14 | 8759.72 (3004.04)<br>2922.86 – 14690.71 | 8669.72 (2832.18)<br>4064.14 – 13758.57 | $p$ = .94<br>$d$ = 0.10 | $p$ = .96<br>$d$ = 0.08 | $p$ = .99<br>$d$ = 0.03 |
| Cognitive variables | | | | | | |
| Visual Memory Accuracy (percent of correct choices)<br>Mean (SD),<br>Range | 63.82<br>(12.64) | 74.70 (11.52) | 71.55 (13.32) | **$p$ = .022**<br>$d$ = 0.90 | $p$ = .10<br>$d$ = 0.60 | $p$ = .66<br>$d$ = 0.25 |
| Visual Memory Response Speed (milliseconds)<br>Mean (SD),<br>Range | 2598.31 (770.08) | 2182.07 (516.35) | 1834.82 (544.64) | $p$ = .089<br>$d$ = 0.63 | **$p$ <.001**<br>$d$ = 1.14 | $p$ = .12<br>$d$ = 0.64 |
| Two-choice Discrimination Accuracy (percent of correct choices)<br>Mean (SD),<br>Range | 81.81<br>(8.44) | 84.71 (5.59) | 88.08 (6.03) | $p$ = .37<br>$d$ = .41 | **$p$ = .007**<br>$d$ = 0.86 | $p$ = .19<br>$d$ = 0.58 |
| Two-choice Discrimination Response Speed (milliseconds)<br>Mean (SD),<br>Range | 1452.16 (414.01) | 1249.08 (209.67) | 1089.39<br>(169.34) | $p$ = .054<br>$d$ = .62 | **$p$ < .001**<br>$d$ = 1.15 | $p$ = .10<br>$d$ = 0.84 |
| Speeded Tapping Response Speed (milliseconds)<br>Mean (SD),<br>Range | 305.27 (199.64) | 178.44 (138.49) | 128.53 (46.64) | **$p$ = .01**<br>$d$ = 0.74 | **$p$ < .001**<br>$d$ = 1.22 | $p$ = .40<br>$d$ = 0.48 |

*Note.* SCOPA-NS = Nighttime sleep disturbance subscale from the SCOPA-SLEEP questionnaire (possible scores range from 0 – 20), SCOPA-DS = daytime sleepiness subscale from the SCOPA-SLEEP questionnaire (possible scores range from 0 – 24),

**Sleep Habits, Physical Activity Levels, and HD-Mobile Cognitive Tasks Performance in Participants With the Expanded HD Repeat**

***Visual Memory Accuracy*.** For visual memory, the first step in the model showed higher DBS was associated with lower visual memory accuracy (β = - 0.46, $p$ = .003). Higher HADS-D scores tended to be associated with lower accuracy, although the relationship did not reach significance (β = - 0.27, $p$ = .074). Together, DBS and HADS-D scores explained a significant amount of the variation in Visual Memory accuracy, $F(2, 34) = 6.77$, $p$ = .03, accounting for 28.5% of the variability in percent correct (adjusted $R^2$ = .24). We added Fitbit time in bed and Fitbit activity calories into the model and observed a significant improvement in the strength of the model, $\Delta F(2, 32) = 3.65$, $p$ = .037, accounting for an additional 13.3% of the variability in Visual Memory accuracy. In the second step of the model, we found participants who spent more time in bed and those with greater DBS tended to have fewer correct responses on the Visual Memory task (the final model is summarised in Table 3). Neither HADS-D scores nor Fitbit activity calories were associated with visual memory accuracy in the final model. Overall, the final model explained 41.8% of the variability in Visual Memory accuracy ($F(4, 32) = 5.74$, $p$ = .001, adjusted $R^2$ = .35).

***Visual Memory response speed.*** The first step of the model, using DBS and HADS-D to predict Visual memory response speed, was not significant $F(2, 33) = 3.14$, $p$ = .057, accounting for 16% of the variability in Visual Memory response speed (adjusted $R^2$ = .11). Neither DBS (β = .31, $p$ = .057) nor HADS-D scores (β = 0.25, $p$ = .12) was not associated significantly with response speed. Nonetheless, it appeared there was a trend towards slower responses when disease burden was higher. Adding Fitbit time in bed and Fitbit activity calories to the model significantly improved the explanation of variance in response speed, $\Delta F(2, 31) = 5.20$, $p$ = .011, accounting for an additional 18.1% of the variability in response speed. In the final model, we found participants who were more physically active (greater

Fitbit activity calories) tended to respond faster (Table 3). The other three predictor variables were not associated significantly with Visual Memory response speed. The final model, $F(4, 31) = 4.57$, $p = .005$, explained 37.1% of the variability in response speed (adjusted $R^2 = .29$).

*Table 3.*

Final multiple regression model associating Visual Memory Task accuracy, Visual Memory Task response time, Two-choice Discrimination accuracy, and Two-choice Discrimination response speed.

| | *B* | *SE B* | β | *p* | $Sr^2$ |
|---|---|---|---|---|---|
| **Visual Memory accuracy** | | | | | |
| **Intercept** | **131.568** | **18.704** | **-** | **< .001*** ** | **-** |
| **DBS** | **- 0.063** | **0.028** | **- 0.327** | **.031*** | **.092** |
| HADS-D | - 0.449 | 0.482 | - 0.135 | .356 | .016 |
| **Fitbit Time in Bed** | **- 0.081** | **0.032** | **- 0.379** | **.018*** | **.114** |
| Fitbit activity calories | 0.003 | 0.004 | 0.094 | .514 | .007 |
| **Visual Memory response speed** | | | | | |
| Intercept | -950.946 | 860.798 | - | .278 | - |
| DBS | 1.519 | 1.284 | 0.177 | .246 | .027 |
| HADS-D | 10.390 | 22.050 | 0.070 | .641 | .004 |
| **Fitbit Time in Bed** | **5.214** | **1.495** | **0.543** | **.001*** ** | **.235** |
| Fitbit activity calories | 0.036 | 0.178 | 0.029 | .842 | .001 |
| **Two-choice accuracy** | | | | | |
| **Intercept** | **85.251** | **12.390** | **-** | **< .001*** ** | **-** |
| DBS | - .016 | 0.018 | - 0.139 | .400 | .017 |
| HADS-D | 0.224 | 0.317 | 0.114 | .485 | .011 |
| Fitbit Time in Bed | - 0.013 | 0.022 | - 0.106 | .536 | .009 |
| **Fitbit activity calories** | **0.007** | **0.003** | **0.445** | **.009** ** | **.178** |
| **Two-choice Response Speed** | | | | | |
| Intercept | 111.390 | 488.432 | - | .821 | - |
| **DBS** | **2.177** | **0.728** | **0.448** | **.005** ** | **.173** |
| HADS-D | 16.401 | 12.511 | 0.195 | .200 | .033 |
| Fitbit Time in Bed | 1.005 | 0.848 | 0.185 | .245 | .027 |
| Fitbit activity calories | - 0.77 | 0.101 | - 0.112 | .452 | .011 |
| **Speeded Tapping Response Speed** | | | | | |
| Intercept | -395.239 | 319.419 | - | .225 | - |
| **DBS** | **1.343** | **0.476** | **0.461** | **.008** ** | **.183** |
| HADS-D | - 7.010 | 8.182 | - 0.139 | .398 | .017 |
| Fitbit Time in Bed | 0.420 | 0.555 | 0.129 | .454 | .013 |
| Fitbit activity calories | 0.005 | - 0.066 | 0.012 | .939 | < .001 |

*Note.* DBS = disease burden score, HADS-D = total score from the depression subscale of the Hospital Anxiety and Depression questionnaire, * $p \leq .05$, ** $p \leq .01$, $p \leq .001$.

***Two-choice Discrimination Accuracy*.** Entering DBS and HADS-D into the first step of the model accounted for 8.9% of the variability in percent correct on Two-choice Discrimination (adjusted $R^2 = .03$), and the model was not significant, $F(2, 33) = 1.62$, $p = .21$, nor were either the DBS or HADS-D, β = - 0.30, $p = .085$, and β = 0.05, $p = .75$, respectively. Entering

Fitbit time in bed and Fitbit activity calories in the second step significantly improved explanatory power, $\Delta F(2, 31) = 4.41$, $p = .021$, accounting for an additional 20.2% of the variability in Discrimination Accuracy. In the final model (4, 31) = 3.187, $p = .027$, which explained 29.1% of variability (adjusted $R^2 = .20$) higher Fitbit activity calories were associated with higher percentages of correct decisions on the task (see Table 3), and no other variables were significantly associated with Discrimination accuracy.

***Two-choice Discrimination Response Speed.*** In the first step of the model, DBS and HADS-D significantly predicted Two-choice Discrimination response speed, $F(2, 33) = 9.11$, $p = .001$, accounting for 35.6% of the variability in response speed. Only DBS was a significant predictor in this model ($\beta = - 0.53$, $p = .001$). The final model, $F(4, 31) = 5.16$, $p = .003$, explained 40% of the variability (adjusted $R^2 = .32$) in Two-choice Discrimination response speed, but adding Fitbit time in bed and Fitbit activity calories to the model did not provide a significant improvement, $\Delta F(2, 31) = 1.14$, $p = .33$, only accounting for an additional 4.4% of the variability in response speed (Table 3).

***Speeded Tapping Task.*** In the first step of the model, DBS ($\beta = - 0.49$, $p = .002$), but not HADS-D ($\beta = - 0.10$, $p = .52$), associated significantly with tapping speed. Together, DBS and HADS-D significantly predicted Speeded Tapping response speed $F(2, 36) = 5.63$, $p = .008$, and explained 25% of the variability (adjusted $R^2 = .21$. The final model, $F(4, 32) = 2.84$, $p = .04$, explained 26% (adjusted $R^2 = .17$) of the variability in tapping speed. Only DBS was a significant predictor; participants with higher DBS had slower tapping speed (Table 3).

## Discussion

The aim of this study was to investigate how sleep and physical activity habits, assessed jointly, related to the cognitive functioning of people with the expanded HD repeat

in naturalistic settings. We observed that, over and above the effects of disease burden and depression symptoms, more time spent in bed and less physical activity associated with poorer functioning on several measures of cognition in our sample of participants with premanifest-and manifest-HD. Specifically, more time in bed associated with reduced accuracy and speed of response on the Visual Memory task. Less physical activity associated with reduced accuracy on the Two-choice Discrimination task, which measures executive functioning and has a working memory component. In contrast, we found no relationship between time in bed or physical activity levels and basic psychomotor speed, as measured by the Speeded Tapping task. To our knowledge, this study provides the first evidence that sleep and physical activity are differentially associated with cognitive function in HD, extending the findings of previous sleep-cognition [15,16] and physical activity-cognition research in HD [17]. Interestingly, we found that time in bed and activity calories did not interact in any significant fashion in any of the regression models. Additionally, through utilising a range of self-report and objective measures of sleep and physical activity, we add to the limited body of knowledge of the day-to-day sleep and physical activity habits of people with the HD-gene expansion. Similar to past research [15,16], bedtime amongst Manifest-HD, Premanifest-HD and Healthy control groups were similar, but compared to the other two groups the Manifest-HD group spent longer in bed and tended to have a later wake time compared with the Healthy Controls. Interestingly, the Manifest-HD group self-reported better sleep quality over the course of the study than the Healthy Control group. We saw no difference between the Premanifest-HD and Healthy Control groups on any of our measures. Overall, physical activity levels amongst the groups appeared similar, which is in keeping with limited previous evidence [17], although individuals in the Manifest-HD group reported a greater time spent sitting.

We expected that physical activity and sleep parameters would provide similar amounts of explanatory power within each regression model, with cognitive task performance relating to both physical activity levels *and* time spent in bed, and that they might interact in some way, but neither of these expectations were met. Moreover, time in bed and physical activity had a weak and non-significant correlation. We found that physical activity and time in bed never jointly associated significantly with the same cognitive functioning measure. For example, greater physical activity levels associated with accuracy of decision making on the Two-choice Discrimination task but had a non-significant relationship to Visual Memory task performance. Conversely, more time in bed associated significantly with Visual Memory, but not Two-choice Discrimination task performance. These findings were counter to our expectations, but actually accord with past studies because, as discussed below, sleep or physical activity appear to affect different domains of cognition to different degrees in HD [15,17,52] and non-HD samples [4,53,54].

Considering exercise and cognition interactions further, studying premanifest-HD participants, Wallace and colleagues[17] found that greater physical activity was associated with better performance on the SDMT, a task with a working memory component, but not with the Stroop Word Reading task, which relies on cognitive inhibition for some trials. Similar to the SDMT, the Two-choice Discrimination task also has a working memory load as participants must keep feedback from previous trials in mind while making decisions on each trial. This is consistent with research with healthy older adults, where a meta-analysis[4] showed that regular physical activity benefits executive control processes, which include working memory. Further research will be required to determine how robust the above findings of physical activity-working memory relationships are in HD and to establish any cause and effect relationship that may exist. This is because in the general population physical activity does not always benefit or associate with executive control processes to a

greater degree than other cognitive domains, such as basic attention or learning and memory [31,55] and physical activity-cognition relationships in HD are likely to be just as complex. Regardless, it appears a pattern is emerging of physical activity associating with cognition in HD with a possibility of some specific association with working memory performance. Little is known about how day-to-day physical activity levels of people with the expanded HD repeat compare to those of the general population, but our findings and those of Wallace and colleagues [17] indicate that general activity levels of people with the expanded HD repeat may be similar to the general population at least until the early manifest stage.

Regarding sleep-cognition interactions, our findings are also consistent with emerging evidence that cognitive functioning in HD is sensitive to sleep habits and timing. For example, Aziz and colleagues [15] and Diago and colleagues [16] found that later wake-time in their samples (which were of similar size to our sample and also contained a mix of manifest- and premanifest-HD participants) was associated with poorer cognitive performance, and we found the same effect on cognitive function with increased time spent in bed. Both later wake time and higher levels of time in bed may reflect impaired night-time sleep. Less time in deep sleep stages, lower duration of rapid eye movement sleep, and higher numbers of awakenings during sleep have been recorded in HD [21,22]. Subsequently, it is reasonable to expect those experiencing such disturbed sleep will require more hours in bed to feel sufficiently rested. It is possible that the effect of time in bed could be confounded with disease stage in our study as the Manifest-HD spent more time in bed and also had the worst cognitive performance. We note, however, that we controlled for disease burden before entering time in bed into the regression models and our results mirror those of previous studies, so we believe our findings are robust.

Contrary to expectation, using the Sleep diary, Manifest-HD participants rated their nightly sleep as *better* than the Healthy Control group. This is surprising as we know from

the literature that people with manifest-HD regularly experience sleep disturbance [21,22] and it is also surprising as there were no between-group differences on the SCOPA-SLEEP night-time sleep scale which measures night-time sleep disturbance. The Manifest-HD group reporting better sleep than Healthy Controls is perhaps related to a combination of related factors including known insight issues in HD [56], the tendency for people with the expanded HD repeat to underreport sleep problems [57], or the subsequent limitations of using self-report sleep questionnaires in HD [22]. Because of the above issues, future studies assessing associations between behaviour, such as sleep and physical activity habits, and cognition, researchers might consider utilizing passive monitoring rather than use of self-report. Coupling ongoing assessment of sleep and physical activity habits with repeated cognitive assessment has the added benefit of in the capturing real time associations between daily behaviours and cognitive functioning in a manner that cannot be achieved with the use of retrospective self-report measures [58-60].

Our study had two primary limitations that must be mentioned. First, our study had a small sample size, which limited the power to detect relationships among variables and the number of predictor variables that could be assessed within the regression models. Regression models can become overfitted when sample sizes are low [61], and so the robustness of the relationships we observed in this study should be assessed in future with larger sample sizes. Second, because Fitbits [47] and actigraphy in general [48] lack accuracy for assessing sleep architecture in HD, we had no means of validating the Manifest HD group's unexpected self-reports of high sleep quality against reliable objective data. Future observational studies of sleep at home in HD could solve this problem utilizing EEG-based devices such as the Dreem Headband, which can be worn in the home and measures aspects of sleep architecture, such as sleep stages, with comparable accuracy to polysomnography [62].

Overall, our results showed that time in bed and physical activity levels have important associations with day-day-cognitive functioning in people with the expanded HD repeat. These findings are congruent with past similar studies in HD [15-17] but extend their findings in two important ways. First, our findings have greater ecological relevance because all assessments were done in the home. Second, we captured 'real time' associations between sleep and physical activity habits and cognitive functioning in a way that was not done in past studies. This is because there was no temporal separation between our ongoing passive monitoring of sleep and physical activity and the assessment of cognition in our study, such as there has been in many in past studies [17,22,57].

Future observational studies of sleep, physical activity, and cognition in HD should also employ longitudinal assessment to determine if, and in what way, long-term physical activity and sleep habits are linked to cognitive symptom progression in HD, an area of knowledge which is sorely lacking. Additionally, given the emerging findings of relationships between sleep timing, physical activity, and cognition in HD, we recommend that these factors be considered as targets of future interventions aimed at improving cognition in HD. The findings of such studies will be critical to allow clinicians to support their patients to manage their symptoms alongside, or even in the absence of, future treatment options that may emerge from the current flux of clinical trials.

References


1. Walker FO. Huntington's disease. *The Lancet.* 2007;369(9557):218-228.
2. Eddy CM, Rickards HE. Impact of cognitive and behavioural changes on quality of life in Huntington's disease. *Basal Ganglia.* 2013;3(2):123-126.
3. Ross CA, Pantelyat A, Kogan J, Brandt J. Determinants of functional disability in Huntington's disease: role of cognitive and motor dysfunction. *Mov Disord.* 2014;29(11):1351-1358.
4. Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults a meta-analytic study. *Psychological science.* 2003;14(2):125-130.
5. Esteban-Cornejo I, Tejero-Gonzalez CM, Sallis JF, Veiga OL. Physical activity and cognition in adolescents: A systematic review. *J Sci Med Sport.* 2015;18(5):534-539.
6. Lautenschlager NT, Almeida OP. Physical activity and cognition in old age. *Current opinion in Psychiatry.* 2006;19(2):190-193.
7. Sibley BA, Etnier JL. The relationship between physical activity and cognition in children: a meta-analysis. *Pediatric exercise science.* 2003;15(3):243-256.
8. Ellenbogen JM. Cognitive benefits of sleep and their loss due to sleep deprivation. *Neurology.* 2005;64(7):E25-E27.
9. Killgore WDS. Effects of sleep deprivation on cognition. *Human Sleep and Cognition: Basic Research.* 2010;185:105.
10. Yaffe K, Falvey CM, Hoang T. Connections between sleep and cognition in older adults. *The Lancet Neurology.* 2014;13(10):1017-1028.
11. Beebe DW. Cognitive, behavioral, and functional consequences of inadequate sleep in children and adolescents. *Pediatric Clinics.* 2011;58(3):649-665.
12. Hamer M, Chida Y. Physical activity and risk of neurodegenerative disease: a systematic review of prospective evidence. *Psychol Med.* 2009;39(1):3-11.

13. Chen J-C, Espeland MA, Brunner RL, et al. Sleep duration, cognitive decline, and dementia risk in older women. *Alzheimer's & Dementia.* 2016;12(1):21-33.
14. Benedict C, Byberg L, Cedernaes J, et al. Self-reported sleep disturbance is associated with Alzheimer's disease risk in men. *Alzheimer's & Dementia.* 2015;11(9):1090-1097.
15. Aziz NA, Anguelova GV, Marinus J, Lammers GJ, Roos RA. Sleep and circadian rhythm alterations correlate with depression and cognitive impairment in Huntington's disease. *Parkinsonism & related disorders.* 2010;16(5):345-350.
16. Diago EB, Martínez-Horta S, Lasaosa SS, et al. Circadian rhythm, cognition, and mood disorders in Huntington's disease. *Journal of Huntington's disease.* 2018;7(2):193-198.
17. Wallace M, Downing N, Lourens S, et al. Is there an association of physical activity with brain volume, behavior, and day-to-day functioning? A cross sectional design in prodromal and early huntington disease. 2016.
18. Vitiello MV. Exercise, sleep, and cognition: interactions in aging. *Exercise and its mediating effects on cognition.* 2008;2.
19. Horne J. Exercise benefits for the aging brain depend on the accompanying cognitive load: insights from sleep electroencephalogram. *Sleep Medicine.* 2013;14(11):1208-1213.
20. Wilckens KA, Erickson KI, Wheeler ME. Physical activity and cognition: a mediating role of efficient sleep. *Behav Sleep Med.* 2018;16(6):569-586.
21. Goodman AO, Barker RA. How vital is sleep in Huntington's disease? *J Neurol.* 2010;257(6):882-897.
22. Herzog-Krzywoszanska R, Krzywoszanski L. Sleep disorders in Huntington's disease. *Frontiers in psychiatry.* 2019;10:221.

23. Pallier PN, Morton AJ. Management of sleep/wake cycles improves cognitive function in a transgenic mouse model of Huntington's disease. *Brain Res.* 2009;1279:90-98.

24. Taylor N, Bramble D. Sleep disturbance and Huntingdon's disease. *The British Journal of Psychiatry.* 1997;171(4):393-393.

25. Lazar AS, Panin F, Goodman AO, et al. Sleep deficits but no metabolic deficits in premanifest H untington's disease. *Ann Neurol.* 2015;78(4):630-648.

26. Petit D, Gagnon J-F, Fantini ML, Ferini-Strambi L, Montplaisir J. Sleep and quantitative EEG in neurodegenerative disorders. *J Psychosom Res.* 2004;56(5):487-496.

27. Arnulf I, Nielsen J, Lohmann E, et al. Rapid eye movement sleep disturbances in Huntington disease. *Arch Neurol.* 2008;65(4):482-488.

28. Hansotia P, Wall R, Berendes J. Sleep disturbances and severity of Huntington's disease. *Neurology.* 1985;35(11):1672-1672.

29. Tabrizi SJ, Reilmann R, Roos RAC, et al. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *The Lancet Neurology.* 2012;11(1):42-53.

30. Ratey JJ, Loehr JE. The positive impact of physical activity on cognition during adulthood: a review of underlying mechanisms, evidence and recommendations. *Rev Neurosci.* 2011;22(2):171-185.

31. Bherer L, Erickson KI, Liu-Ambrose T. A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. *Journal of aging research.* 2013;2013.

32. Fritz NE, Rao AK, Kegelmeyer D, et al. Physical Therapy and Exercise Interventions in Huntington's Disease: A Mixed Methods Systematic Review. *Journal of Huntington's Disease.* 2017;6:217-235.

33. McLaren B, Andrews SC, Glikmann-Johnston Y, et al. Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *J Neurol.* 2020.

34. Group H. Unified Huntington's disease rating scale: reliability and consistency. *Mov Disord.* 1996;11(2):136-142.

35. Zhang L, Zhao Z-X. Objective and subjective measures for sleep disorders. *Neuroscience Bulletin.* 2007;23(4):236-240.

36. Penney JB, Vonsattel JP, Macdonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol.* 1997;41(5):689-692.

37. Zhang Y, Long JD, Mills JA, et al. Indexing disease progression at study entry with individuals at-risk for Huntington disease. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics.* 2011;156B(7):751-763.

38. Unified Huntington's disease rating scale: Reliability and consistency. *Mov Disord.* 1996;11(2):136-142.

39. McLaren B, Andrews SC, Glikmann-Johnston Y, et al. Feasibility and initial validation of 'HD-Mobile', a smartphone application for remote self-administration of performance-based cognitive measures in Huntington's disease. *J Neurol.* 2020.

40. Marinus J, Visser M, van Hilten JJ, Lammers GJ, Stiggelbout AM. Assessment of Sleep and Sleepiness in Parkinson Disease. *Sleep.* 2003;26(8):1049-1054.

41. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361-370.

42. Cosco TD, Doyle F, Ward M, McGee H. Latent structure of the Hospital Anxiety And Depression Scale: a 10-year systematic review. *J Psychosom Res.* 2012;72(3):180-184.

43. De Souza J, Jones LA, Rickards H. Validation of self-report depression rating scales in Huntington's disease. *Mov Disord.* 2010;25(1):91-96.

44. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity.* 2015;12(1):42.

45. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. *Sleep and Breathing.* 2012;16(3):913-917.

46. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport.* 2014;17(5):496-500.

47. Maskevich S, Jumabhoy R, Dao PD, Stout JC, Drummond S. Pilot Validation of Ambulatory Activity Monitors for Sleep Measurement in Huntington's Disease Gene Carriers. *Journal of Huntington's Disease*. 2017;6(3):249-253.

48. Townhill J, Hughes AC, Thomas B, et al. Using Actiwatch to monitor circadian rhythm disturbance in Huntington' disease: A cautionary note. *J Neurosci Methods.* 2016;265:13-18.

49. Walters EM, Phillips AJ, Mellor A, et al. Sleep and wake are shared and transmitted between individuals with insomnia and their bed-sharing partners. *Sleep.* 2020;43(1):zsz206.

50. Straus LD, Drummond SP, Nappi CM, Jenkins MM, Norman SB. Sleep variability in military-related PTSD: A comparison to primary insomnia and healthy controls. *J Trauma Stress.* 2015;28(1):8-16.

51. Hamill K, Jumabhoy R, Kahawage P, de Zambotti M, Walters EM, Drummond SPA. Validity, potential clinical utility and comparison of a consumer activity tracker and a research-grade activity tracker in insomnia disorder II: Outside the laboratory. *J Sleep Res.* 2020;29(1):e12944.

52. Diago EB, Pérez JP, Lasaosa SS, et al. Circadian rhythm and autonomic dysfunction in presymptomatic and early Huntington's disease. *Parkinsonism & related disorders.* 2017;44:95-100.

53. Motl RW, Gappmaier E, Nelson K, Benedict RH. Physical activity and cognitive function in multiple sclerosis. *Journal of Sport and Exercise Psychology*. 2011;33(5):734-741.

54. Eggermont LH, Milberg WP, Lipsitz LA, Scherder EJ, Leveille SG. Physical activity and executive function in aging: the MOBILIZE Boston Study. *J Am Geriatr Soc.* 2009;57(10):1750-1756.

55. Smith PJ, Blumenthal JA, Hoffman BM, et al. Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials. *Psychosom Med.* 2010;72(3):239-252.

56. Ho AK, Robbins AO, Barker RA. Huntington's disease patients have selective problems with insight. *Mov Disord.* 2006;21(3):385-389.

57. Zhang Y, Ren R, Yang L, et al. Sleep in Huntington's disease: a systematic review and meta-analysis of polysomongraphic findings. *Sleep.* 2019;42(10):zsz154.

58. Allard M, Husky M, Catheline G, et al. Mobile technologies in the early detection of cognitive decline. *PLoS ONE.* 2014;9(12):e112197.

59. Kaye JA, Maxwell SA, Mattek N, et al. Intelligent Systems for Assessing Aging Changes: Home-Based, Unobtrusive, and Continuous Assessment of Aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences.* 2011;66B(suppl 1):i180-i190.

60. McCrae CS, Vatthauer KE, Dzierzewski JM, Marsiske M. Habitual sleep, reasoning, and processing speed in older adults with sleep complaints. *Cognitive therapy and research.* 2012;36(2):156-164.

61. Tabachnick BG, Fidell LS. Using multivariate statistics. 2007.

62. Arnal PJ, Thorey V, Ballard ME, et al. The Dreem Headband as an Alternative to Polysomnography for EEG Signal Acquisition and Sleep Staging. *bioRxiv.* 2019:662734.

## Chapter Six: General Discussion

The overarching goal for my thesis research was to design, create, test, and then utilise an exclusively remote means of collecting clinically and scientifically useful data on the day to day cognitive functioning and sleep and physical activity habits of people with premanifest- and early manifest-HD. This involved demonstrating that people with the expanded HD repeat could complete the protocol (i.e., feasibility, acceptability, which was covered in Chapter Two) and that the HD-Mobile cognitive tasks would be reliable, achieve several forms of validity, and be sensitive to cognitive functioning in HD compared to controls. These aims were addressed in Chapters Three and Four. I also had the goal of adding to the limited literature on how sleep and physical activity habits relate to cognitive functioning in HD. This aim was addressed in Chapter Five. I will next present a summary of the key background information that was related to my thesis research, which I presented in Chapter One, and will also summarise the main findings from the empirical research that was presented in Chapters Two to Five.

In Chapter One I explained that the work presented in my thesis was prompted by three realisations: 1) The field of Huntington's disease study needed a rapid, inexpensive, accurate, and convenient means of collecting cognitive data in order to provide rich datasets that can be used to aid the identification of modifiers of cognitive symptom onset and expression; 2) Sleep and physical activity were understudied as factors potentially influencing cognition in HD, particularly in the context of home-based monitoring where cognition, sleep, and physical activity outcomes could be measured in combination rather than using retrospective reports; and 3) Remote assessment using mobile apps is a methodology that has sufficiently matured in recent years to make it possible for people from the HD-community to perform cognitive tests and questionnaires via their smartphones.

Lastly, I explained that no one had attempted to combine remote assessment of cognition, sleep, and physical activity in HD. The above points all lead into my listed research aims.

In Chapter Two, I demonstrated the feasibility and acceptability of the remote assessment protocol. Participants with the expanded HD repeat and control participants completed the protocol with equally high levels of confidence and did so without in-person training. In Chapter Three, I showed that the HD-Mobile cognitive tasks had robust test-retest reliability, known groups validity, concurrent validity, and sensitivity to indicators of disease progression. Next, in the study presented in Chapter Four, I successfully applied cognitive modeling to data from the Two-choice Discrimination task and detected group-level differences in evidence accumulation speed between premanifest- and manifest-HD and control participants which were not apparent when viewing the observable performance on the task. Lastly, in Chapter Five, I showed that reduced physical activity and more time spent in bed was associated with poorer cognitive functioning in the combined group of premanifest- and manifest-HD participants. Together, these findings demonstrate that remote assessment of cognition, sleep, and physical activity in HD, in which participants are asked to self-administer all assessments without in-person training, is feasible and acceptable, and yields data that is reliable and achieves several forms of validity. The findings also showed that the remote assessment methodology produced useful information about the day to day cognitive functioning of people with the expanded HD repeat, including how their cognitive functioning relates to their sleep and physical activity habits.

## 6.1. Integrated Overview of Findings and Implications

**6.1.1. Remote assessment of cognition, sleep, and physical activity in HD is feasible, reliable, valid, and provides scientifically and clinically meaningful Data.** The first major implication of my findings is that remote, smartphone-based, cognitive assessment in HD is feasible *in the absence of providing in-person training*. This has rarely, if at all, been achieved

with a cognitively compromised sample and has important implications for the scalability of cognitive assessment in HD and other cognitively compromised populations. First, regarding novelty, we know of no previous studies which have attempted the same approach. Koo and Vizer (2019) reviewed studies that used mobile phone and tablet based cognitive assessment tools in older adults. They identified 29 studies in total, of which only five included any home-based assessment. Of those five studies, only one (Tung et al., 2014) included a cognitively impaired sample, however, in that study participants did not self-administer any cognitive tasks, but rather only wore GPS trackers. A prior review paper by Zygouris and Tsolaki (2015) summarised computerised cognitive assessment of older adults more broadly, including a review of all computerised batteries, not just those based on smartphones or tablets. They identified a total of 17 batteries, some of which involved self-assessment, but in all cases, participants were either trained how to use the assessment battery, or they were supervised by a clinician. We found one study that, similar to our methodology, utilised self-administration of cognitive tasks on smartphones in the home environment (Lipsmeier et al., 2018), however, unlike our study, the participants underwent in-person training at the beginning of the study. Thus, my thesis research project has made an important and novel advancement in cognitive assessment by demonstrating that people with HD can self-administer cognitive tasks in their home, in the absence of attending the clinic for training at any stage.

Smartphones are now ubiquitous, and the ability to employ smartphones in the absence of any in-person training provides major benefits by limiting the need for patients and research participants to attend hospitals and clinics (Bauer et al., 2012; Klimova, 2017; Koo & Vizer, 2019; Rentz et al., 2016). Indeed, mobile devices have the potential to provide ongoing, repeated, cognitive assessment, or even just cognitive screening, without requiring patients to attend the clinic, and reduce associated staffing and resource requirements in comparison to in-person assessment. These have been primary reasons for the development of mobile testing

approaches (Darby et al., 2014; Klimova, 2017; Koo & Vizer, 2019; Rentz et al., 2016). The benefits of home-based testing also accrue in that many cognitively impaired people lack independent transport options to get to appointments, and this is particularly important in people with HD who often lose the ability to drive (Beglinger et al., 2010; Farrell et al., 2019; Vaccarino et al., 2011). The benefits of not having to attend a clinic for cognitive assessment extends to people who live in geographically remote areas. The outbreak and global pandemic of coronavirus 2019 (COVID-19) also underscores the utility of having remote assessment options for situations where the extreme infectiousness of a virus makes face-to-face assessments dangerous (Yuan, Li, Lv, & Lu, 2020). Hence, our methodology, which combines mobile assessment with self-administration, fully maximises the benefits of mobile based assessment.

The second important outcome of my research is that HD-Mobile can reliably assess several aspects of cognitive functioning, including psychomotor speed (e.g., Speeded Tapping task), visual learning and memory (e.g., the OILL task), and several aspects of decision making (e.g., the Two-choice Discrimination task). Because we had to address many unknowns at the start of this study, we selected tasks that we thought would work in an unsupervised setting with only minimal onscreen instructions, and deliberately limited the duration of the cognitive battery. Nonetheless, there is the potential to add more or different tasks in future studies. Demonstrating reliability and validity of any assessment tools should be an essential requirement before they can be used in any context, followed by ongoing validation efforts (American Psychological Association, 2017; Bauer et al., 2012; Daniëls et al., 2020). The results from my thesis research showed that HD-Mobile is capable of detecting aspects of cognitive functioning that are relevant to HD, which is an essential property for a battery to have if it is, for example, to be used for assessing the outcomes of a clinical trials (Stout, Andrews, & Glikmann-Johnston, 2017). Therefore, in future applications, HD-Mobile and the

cognitive tasks within could be a useful adjunct within clinical trials and for post-marketing monitoring of treatments, either in the immediate sense and whether or not functioning in these areas remain stable following treatment. Moreover, because the mobile-based platform allows more frequent assessment than clinic-based assessment, both variability in cognitive functioning and how trial participants function in the more naturalistic environment of their own homes, outside of the clinic setting, can be assessed in a way not possible with clinic visits, which usually occur relatively infrequently (Allard et al., 2014; Kaye et al., 2011; Koo & Vizer, 2019; Moore et al., 2017; Toh et al., 2014).

My thesis results also demonstrate that we were able to reliably measure the sleep and physical activity habits of people with the expanded HD repeat, and that we could detect relationships between sleep, physical activity, and cognitive functioning. This result is particularly important for understanding if, and to what degree, there is interplay between these factors and HD signs and symptoms. The major benefit of coupling ongoing assessment of contextual, or environmental, factors such as sleep and physical activity habits with repeated cognitive assessment in the home is that real time associations between daily behaviours and cognitive functioning can be captured (Allard et al., 2014; Kaye et al., 2011; McCrae et al., 2012). Moreover, prospective objective assessment, in contrast to asking for retrospective self-reports of habits and behaviour, is generally more accurate and reliable (Allard et al., 2014; McCrae et al., 2012). Benefits of this approach are particularly important in HD research because limited insight and underreporting of sleep difficulties are known to occur in retrospective self-reports of people with HD (Ho, Robbins, & Barker, 2006; Hoth et al., 2007; Videnovic et al., 2009; Y. Zhang et al., 2019). Better understanding of how environmental factors relate to symptom progression and onset in HD is critical for the development of non-pharmacological interventions, physical activity interventions, or efforts to improve sleep

habits and sleep hygiene (Fritz et al., 2017; Goodman & Barker, 2010; Mo, Hannan, & Renoir, 2015), for example, but has not been achieved as of yet.

One of the best known studies of the relationship between lifestyle factors and motor symptom onset in HD relied entirely on retrospective self-report measures to capture information on lifestyle (Trembath et al., 2010). In this study participants had to retrospectively recall their engagement in leisure-time physical and intellectual activities during each decade of their life from their teens until their diagnosis. Thus, the accuracy of these results is dependent on the accuracy of each person's recall of their activity levels across multiple decades of life. Results indicated that passive lifestyles (i.e., lower levels of physical activity and less cognitively engaging activities) were associated with earlier motor onset. Although it would be impractical to conduct a mobile app study across such a long period as decades, prospective assessment in large samples, longitudinally, over even 24 months, may help to create a more accurate picture of the long term levels of physical activity in people with HD. Our method of monitoring also provides a means to improve understanding of how lifestyle factors, such as sleep and exercise habits, relate to cognitive symptom progression via information-rich prospective observational studies. Importantly, because all assessments can be done in the home, with high frequency, the cognitive data obtained will have higher ecological validity than data collected at infrequent time points in the lab or in the clinic.

A final important implication of my research is that because we have shown that remote cognition, sleep, and physical activity can be assessed in HD, without meeting with participants or providing them with formal training, we can be confident that our approach can also be applied to other neurodegenerative or cognitively compromised populations. We have already seen that people with early signs of dementia and people with Parkinson's disease are able to self-administer cognitive and other tests on smartphones and computer devices after in-person training (Koo & Vizer, 2019; Lipsmeier et al., 2018; Zygouris & Tsolaki, 2015). The results

from my thesis project show that with a careful and considered approach, fully remote assessment, without in-person training, and with all the associated benefits, should be possible in other neurodegenerative populations. We believe that an important key to our success was the community consultation that we undertook in the process of developing the project and the methodology; we went through several rounds of input and idea exchange, and review of our app as it developed, with a sample of just those people who we would ask to use the app on their own; I address the importance of community consultation further in the section that follows. Next, I discuss some key clinical implications of findings from my thesis research.

**6.1.2. Clinical implications of the thesis findings.** In Chapter Four I demonstrated that both premanifest- and manifest-HD participants had less evidence of accumulation speed compared to controls, differences that only emerged following diffusion modeling of the two-choice discrimination task. Importantly, in the absence of applying mathematical modeling to the Two-choice Discrimination task data, differences in performance between premanifest-HD and control groups were undetectable. The finding from diffusion modelling was that premanifest- and manifest-HD participants had a reduced rate of evidence accumulation, a manifestation of slowed processing, which has functional implications for contexts in which processing of visual evidence is required for making decisions. Driving is one such context (Daneshi, Azarnoush, & Towhidkhah, 2020).

Driving is a complex functional ability, which is commonly impaired in HD (Farrell et al., 2019; Jacobs, Hart, & Roos, 2017; Williams, Downing, Vaccarino, Guttman, & Paulsen, 2011). Several studies have had some success using cognitive batteries to predict fitness to drive in patients with manifest-HD, particularly using processing speed, cognitive flexibility, and visual attentional control as predictors (Beglinger et al., 2012; Devos, Nieuwboer, Tant, De Weerdt, & Vandenberghe, 2012; Farrell et al., 2019; Hennig, Kaplan, Nowicki, Barclay, & Gertsberg, 2014; Jacobs et al., 2017). Nonetheless, no consensus or gold-standard cognitive

battery exists for predicting fitness to drive in HD (Jacobs et al., 2017) and cognitive batteries have shown varied predictive utility for driving fitness in HD (Farrell et al., 2019; Jacobs et al., 2017). Diffusion modeling of two-choice discrimination tasks, such as the one used in my thesis, may make a valuable addition to any cognitive battery aimed at predicting driving fitness in HD. For example, the ability to estimate the time to collision of two objects is a critical ability when driving (Daneshi et al., 2020) and diffusion models have been shown be the best choice for modeling data from time to collision studies (Daneshi, Azarnoush, Towhidkhah, Gohari, & Ghazizadeh, 2019). Indeed, Daneshi et al. (2020) showed, in a driving simulator task that assessed time to collision decisions, that healthy participants increased their evidence accumulation speed when required to make driving decisions under time pressure compared to when not under time pressure. Relevant to these findings, in Chapter Four, I showed that premanifest- and manifest-HD participants were unable to accelerate evidence accumulation in the Two-choice Discrimination task to the same extent as healthy controls. Inability to accelerate evidence accumulation under time pressure when driving could have disastrous consequences for people with premanifest- and manifest-HD as it may heighten propensity for accidents. Hence, the evidence accumulation deficits we observed in Chapter Four may have implications for driving ability, and as such, tasks assessing evidence accumulation ability may have promise as predictors of driving ability in HD.

Another clinical implication of the modelling results from Chapter Four relates to the possible neurological correlates of the reduced evidence accumulation speed in HD participants. The slowed evidence accumulation may be due to structural deficits in the fronto-parietal pathway since the fronto-parietal pathway has been linked with evidence accumulation speed in the general population (Brosnan et al., 2020; Mulder, Van Maanen, & Forstmann, 2014), and because the pathway is known to be affected in manifest-HD (Poudel et al., 2014). Poudel et al. (2014) demonstrated that higher radial diffusivity, indicative of demyelinative

processes (Song et al., 2005), correlated with poorer performance on SDMT and Stroop and with worse motor symptoms in manifest-HD participants. Even though Poudel et al. (2014) observed significantly greater radial diffusivity in premanifest-HD versus controls, these levels did not associate significantly with motor signs or any cognitive measures. Hence, we add evidence of another specific cognitive deficit that may be related to neuronal changes in the fronto-parietal pathway in HD, that of reduced evidence accumulation speed. Moreover, we show that these deficits have functional implications in premanifest-HD, not just manifest-HD. For future studies the role of the fronto-parietal pathway, and that of any other pathways, in evidence accumulation in HD can be substantiated with research that integrates cognitive modelling with neuroimaging.

Findings from the thesis project on the relationships between sleep, physical activity, and cognition in HD also have important clinical implications. In Chapter Five, I showed that greater time spent in bed and lower levels of physical activity, averaged over seven days, were related to poorer cognitive functioning on HD-Mobile cognitive tasks, measured on the eighth day. Both findings were congruent with past findings in HD which showed that later habitual wake times (Aziz et al., 2010; Diago et al., 2018) and lower levels of physical activity (Wallace et al., 2016) related to poorer cognitive functioning. Our findings added a new component, ecological validity, in that all data in my studies were collected within the home environment, meaning we had a direct measure of how sleep and physical activity habits relate to the day-to-day functioning of people with the expanded HD repeat rather than how they function under the specialised settings of a clinic or laboratory. Moreover, in my research cognitive assessments were conducted in close proximity to the measures of sleep and physical activity, which is key for understanding the dynamic interaction of these factors (McCrae et al., 2012; Rentz et al., 2016). Our results also prompt the need for further sleep and physical activity interventions, and longitudinal observational studies in HD so that cause and effect

relationships between sleep, physical activity, cognition, and broader HD signs and symptoms, can be confidently established. Indeed, careful investigation of environmental and behavioural modifiers of cognitive and other symptoms in HD are sorely needed (Fritz et al., 2017; Goodman & Barker, 2010; Mo et al., 2015). Additionally, our results suggest that sleep and physical activity interventions that aim to improve cognition should consider managing physical activity levels and sleep timing, particularly managing time in bed and waking time.

**6.1.3. Community consultation in the design and development of research programs.** In Chapter Two, I described how we involved members of the HD community in the development of the study protocol. The strategy of meeting with members of the HD community was guided, in part, by the United States Food and Drug Administration's Voice of the Patient reports (US Food and Drug Administration, 2017). The reports advocate for involving patients in the process of designing and running research projects, courses of treatments, and the development of guidelines that affect the patients. Indeed, collaboration between researchers, clinicians, and patients is an essential component of good research practice and for the successful design and implementation of healthcare (Coulter & Ellins, 2006; Leidy & Vernon, 2008; Ocloo & Matthews, 2016; Payne et al., 2011). In particular, collaboration and consultation with patients and participants is critical when researchers plan to ask people with cognitive impairments to self-administer cognitive tests (Hagelkruys et al., 2016). For example, Leidy and Vernon (2008) recommended that involvement of participants during the development of new instruments include small scale focus groups during development, followed by one-one cognitive debriefing sessions. In the cognitive debriefing sessions, participants can provide further specific feedback about the instrument, such as if they had difficulty understanding instructions, with feedback used to further improve the assessment tool (Leidy & Vernon, 2008). The benefit of this process was shown in Chapter Two. Our lab specialises in HD research and our research team collectively has multiple decades of HD

research experience, yet the HD community members we consulted with had several insights and suggestions for protocol design that we had not considered. For example, we planned to include the International Physical Activity Questionnaire (Hagströmer et al., 2006) within HD-Mobile, but decided not to because the HD-community members found the questions too long and confusing. We made further adjustments to our protocol following the "cognitive debriefing" sessions, such as adding a clear flow chart of required activities to the participant instructions, so that participants could easily understand what they were being asked to do. The input from community members likely contributed to there being zero attrition amongst the 91 participants who took part in my research. Although, the short protocol length was also likely a factor. These results further demonstrate the importance of patient consultation in the development of research, and I argue that further patient-centered focus will be crucial for the development of longitudinal research protocols, which is the planned next step following from my thesis.

Poor adherence is an obstacle for remotely completed longitudinal research protocols and for attempts at phone-based interventions (e.g., Darby et al., 2014; Henson, Peck, & Torous, 2019; Jongstra et al., 2017b; Lipsmeier et al., 2018). For example, when requiring people with Parkinson's disease to complete brief daily tasks on a smartphone, Lipsmeier et al. (2018) had an adherence rate of 61% at the six month mark. Similarly, Darby et al. (2014), asked healthy older adults to complete a web-based cognitive battery once a month over a twelve month period. 95% percent of participants completed the battery in the first month, but this fell to 67% at month three and 43% at month twelve. The above studies do not report on whether or not the design period involved consultation with consumers. Nonetheless, it is clear that if remote, self-administered, cognitive assessment strategies are to be useful adjuncts to clinical trials or useful for capturing environmental modifiers of functioning, as proposed (Allard et al., 2014; Koo & Vizer, 2019; Rentz et al., 2016), then better adherence is needed.

This is especially true as people that remain in remote assessment studies when others drop out may not be truly representative of the population from which they were drawn, reducing the generalisability of study results (Darby et al., 2014). Careful consultation with patients and their families, which we utilised in the study we report in Chapter Two, and as is advocated for by many (Coulter & Ellins, 2006; Hagelkruys et al., 2016; Leidy & Vernon, 2008; US Food and Drug Administration, 2017), and a focus on development of the therapeutic alliance between researchers and research participants, as advocated for by Henson, Peck, et al. (2019), may help to improve adherence in longitudinal remote assessment studies. In the study presented in Chapter Two we used the same protocol development process outlined by Leidy and Vernon (2008) for facilitating patient involvement in research and medicine. Specifically, we met with a group of community members during protocol development to understand their needs and preferences, and then met one on one with participants after they completed the protocol to get their experience-driven advice on how to further improve the protocol. An identical process may help to identify and eliminate reasons for study dropout in longitudinal versions of our protocol. Regarding therapeutic alliance, Henson, Peck, et al. (2019) and Henson, Wisniewski, Hollis, Keshavan, and Torous (2019) have argued that a focus on development and maintenance of the therapeutic alliance between clinicians or researchers and participants has often been overlooked by developers of digital mental health intervention tools, and has resulted in poor uptake of and adherence to digital mental health tools. Therapeutic alliance relates to the relationship and rapport between patient and therapist and a working alliance between the two, including an agreement on goals and tasks to be completed as well as the development of a bond between patient and therapist (Bordin, 1979; Henson, Peck, et al., 2019). Henson, Peck, et al. (2019) propose that adherence and engagement with mental health intervention apps will be improved if therapists liaise with patients and set goals related to app usage. A similar process may be beneficial in longitudinal remote assessment, whereby

clinicians and researchers regularly have contact with patients or participants and set and track goals related to test administration. This would have staffing requirements that would to some degree undo one of the proposed benefits of remote self-administration (Darby et al., 2014; Klimova, 2017). Nonetheless, if adherence rates were substantially improved, the benefits to quality and amount of data collected may outweigh the additional staffing costs.

### 6.2. Limitations and Future Directions

There were three primary limitations in my project that were related to our use of Fitbit One devices in the protocol. The first of these was a smaller than ideal sample size, the second was some loss of sleep and exercise data due to improper use of Fitbits, and the third was the limited ability of the Fitbit to directly measure sleep in HD. Each of these lead to options for future directions. Regarding sample size, our small sample size was a general limitation across Chapters Three to Five, limiting the power of some statistical analyses. Sample size was small because our recruitment was limited by logistical issues tied to the use of a set number of Fitbits available for our study. Rate of progress in collecting data was also limited by the time it took for the entire process of posting Fitbits to participants, having them complete the protocol, and then send the Fitbits back. This raises the first possible future direction, which involves further development of HD-Mobile to remove the need for the use of third-party activity trackers such as the Fitbit. Smartphone applications can now be used to collect large amounts of information from the phones themselves, including basic activity information such steps taken and distance travelled, sit to stand transitions, aspects of sleep timing, as well as details of phone usage such as the amount of messages sent, time taken to reply to messages, and other similar details (Barnett et al., 2018; Ben-Zeev, Scherer, Wang, Xie, & Campbell, 2015; Case, Burwick, Volpp, & Patel, 2015; Henson, Wisniewski, et al., 2019; Lipsmeier et al., 2018; Torous, Kiang, Lorme, & Onnela, 2016). Hence, utilising the wide range of data collected by smartphones presents a valid option for passively collecting data on activity levels and other aspects of functioning,

without the requirement for third party monitoring devices, potentially limiting the chances of operator error and lost data. Alternatively, future protocols can continue to utilize third party trackers, which continue to improve in terms of user friendliness, the accuracy of the data they generate on physical activity, and the diversity of the data outcomes they provide, such as heart rate measures, and more accurate and in depth measurement of sleep (Arnal et al., 2019; Boe et al., 2019; Kang, Kim, Byun, Suk, & Lee, 2019; Thomson et al., 2019). Hence, if third party trackers are used, a greater depth and quality of information can be collected and likely with a lesser chance of loss of data due to operator error. The choice of method for future studies will depend on the type and amount of data needed and the associated costs. For example, third party devices may be useful for collecting data on heart rate which cannot be collected by a smartphone alone, but the cost of providing wearable devices to a large cohort of research participants may be prohibitive.

Both Fitbit One devices, and actigraphy based sleep monitoring devices are limited in their ability to assess aspects of sleep architecture such as total sleep time, sleep efficiency, number of awakenings, etc, in HD (Maskevitch et al., 2017; Townhill et al., 2016). Hence, we could only use the Fitbit to measure overall sleep/wake patterns. This represents the third limitation of using Fitbit devices in the study because we were unable to validly assess the relationship between sleep architecture and cognitive functioning in participants with the expanded HD repeat. Although the results from Chapter Five and the studies by Aziz et al., 2010 and Diago et al., 2018 demonstrate that sleep timing, particular time in bed and habitual wake times have important relationships with cognitive functioning, the ability to precisely assess sleep architecture in the home will likely be of great benefit for developing a fine detailed understanding of how sleep functioning and cognitive functioning relate in HD in a community setting. Subsequently this knowledge may assist the development of targeted sleep interventions in HD. New technologies using headbands that provide electroencephalogram

recordings and can also measure heart rate, breathing rate, eye movements, and other key sleep indicators (Arnal et al., 2019; Shambroom, Fabregas, & Johnstone, 2012) have promising utility, as compared to lab-based polysomnography, for assessing elements of sleep architecture in healthy individuals in home and community settings. These devices appear to be promising for assessing sleep architecture/cognition relationships in HD. However, before conducting a large-scale study investigating relationships between sleep variables and cognition using any of these methods, they should first be specifically validated against polysomnography with a sample of premanifest and especially manifest HD participants to understand how validly the devices measure sleep in HD.

A final future direction for the research presented in this thesis is further development and implementation of two-choice decision making tasks and cognitive modelling in HD. I have argued above that two-choice decision making task data, after processing with diffusion models, may be a particularly useful addition to cognitive batteries aimed at predicting fitness to drive in HD, and this relationship could be further examined in future studies. Our group published the first papers showing decision making impairments in HD nearly 20 years ago (Stout et al., 2001) and followed this work with cognitive modelling that disentangled the cognitive processes that were impaired in HD compared to healthy controls (Busemeyer & Stout, 2002). The earlier work from the Stout lab used a more complex four-choice decision making task, which would be unsuitable for unsupervised home-based mobile assessment, however, and used a different class of cognitive models. That approach too helped to uncover the cognitive processes affected in HD that underlay their decision-making performance, in particular, the tendency to overweigh the most recent outcomes of their choices, and the tendency to learn relatively better from positive than negative feedback. Overall, however, the use of formal models in studies of cognition in neurodegenerative disease is rare, despite the evidence that they can generate sensitive and clinically relevant outcomes that enhance what

can be learned from performance-based measures of cognitive functions. Parameters generated by cognitive models of task performance have also been linked to specific known neural correlates. For example, response cautiousness, which is instantiated in the *boundary* parameter in the diffusion model we used (Voss et al., 2013) is strongly linked to the function of the striatum (Bogacz et al., 2010; Herz et al., 2018; Damian M. Herz, Zavala, Bogacz, & Brown, 2016), which is a primary site of neuronal atrophy in HD (F. O. Walker, 2007). If cognitive modelling can be developed to be efficiently implemented in the future, and then be applied in larger scale studies in neurodegenerative samples using remotely collected densely sampled data, it will be possible to herald an era in cognitive neuroscience in which 'big data' approaches can be applied to push frontiers of knowledge about cognition in neurodegenerative diseases, and its role in clinical progression.

Overall, my research supports the value of Two-Choice Discrimination Tasks in further research, but as I note in Chapter Four, future studies would benefit from task conditions in which speed accuracy trade-offs are deliberately manipulated. Such investigations may show how well people with HD can shift their response style based on the demands of the task context, which is a key capability for many aspects of daily life, such as driving. Regardless of outcome, such an approach should provide new valuable information about decision making processes in premanifest- and manifest-HD and how these might affect daily decision-making, particularly impulsivity, and functioning across the HD disease process.

Future studies incorporating neuroimaging and diffusion modelling in HD would be beneficial, although obviously imaging in the context of at-home mobile assessment is impractical. Separate aspects of data collection however, using fMRI for example (see Mulder et al., 2014 for a review of how this has been conducted in the general population), could facilitate an understanding of how subcortical pathology in HD relates to decision making processes, such as speed of encoding and regulation of decision making thresholds. If decision

variables are found to link closely to imaging measures, this heralds the possibility that decision making tasks could then be used as markers of disease progression in people with the expanded HD repeat. Further, given treatments for HD are aimed at interrupting subcortical atrophy, decision making tasks may be potential proxies for monitoring response to therapeutics. Clearly, the implementation of such an approach requires significant additional developments, but the research in this thesis provides some initial evidence that invites speculation about how cognitive testing with modelling might be developed for future studies and even clinical trials.

### 6.3. Final Conclusions

The goal for this thesis was to design, create, test, and then utilise an exclusively remote means of collecting clinically and scientifically useful data on the day to day cognitive functioning and sleep and physical activity habits of people with premanifest and early manifest HD. To my knowledge, this was the first project to attempt fully remote cognitive assessment in a neurodegenerative population where the participants self-administered all tasks without receiving any in-person training. To my knowledge my project was also the first to apply diffusion modelling to data collected from smartphones and the first to investigate the relationships between sleep and physical activity habits and cognitive functioning in HD, with all data collected in participants' homes. Overall, we provided several novel findings to the field of HD research.

To summarise the key findings, I showed that our methodology for remote assessment of cognition, sleep, and physical activity in HD was feasible and acceptable by people with the expanded HD repeat and that the data collected was reliable, valid, and scientifically meaningful. Importantly, the cognitive data collected was sensitive to cognitive functioning in HD. We were able to separate manifest-HD from premanifest participants and controls on cognitive performance, and also to separate premanifest-HD from controls on cognitive performance. Additionally, we were able to assess and find important relationships between

the sleep and physical activity habits of participants and their cognitive functioning. All of these findings were achieved without ever needing to meet participants, without them needing to travel to a clinic or laboratory and without the need to provide any formal training. I hope the findings of this thesis will prompt a new approach to assessment of cognition and important lifestyle factors in HD that speed the discovery of modifiers of disease onset and improve the assessment of therapeutic outcomes, hence benefiting the HD community.

General references

Allard, M., Husky, M., Catheline, G., Pelletier, A., Dilharreguy, B., Amieva, H., . . . Swendsen, J. (2014). Mobile technologies in the early detection of cognitive decline. *PLoS ONE, 9*(12), e112197.

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. Retrieved from https://www.apa.org/ethics/code/

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. (2003). The role of actigraphy in the study of sleep and circadian rhythms. American Academy of Sleep Medicine Review Paper. *Sleep, 26*(3), 342-392.

Antoniadis, E., Ko, C., Ralph, M., & McDonald, R. (2000). Circadian rhythms, aging and memory. *Behavioural Brain Research, 111*(1), 25-37.

Arnal, P. J., Thorey, V., Ballard, M. E., Hernandez, A. B., Guillot, A., Jourde, H., . . . Chennaoui, M. (2019). The Dreem Headband as an Alternative to Polysomnography for EEG Signal Acquisition and Sleep Staging. *bioRxiv*, 662734.

Arnulf, I., Nielsen, J., Lohmann, E., Schieffer, J., Wild, E., Jennum, P., . . . Tabrizi, S. (2008). Rapid eye movement sleep disturbances in Huntington disease. *Archives of Neurology, 65*(4), 482-488.

Austin, M.-P., Mitchell, P., & Goodwin, G. M. (2001). Cognitive deficits in depression: Possible implications for functional neuropathology. *British Journal of Psychiatry, 178*(3), 200-206. doi:10.1192/bjp.178.3.200

Aziz, N. A., Anguelova, G. V., Marinus, J., Lammers, G. J., & Roos, R. A. (2010). Sleep and circadian rhythm alterations correlate with depression and cognitive impairment in Huntington's disease. *Parkinsonism & related disorders, 16*(5), 345-350.

Baker, C. R., Stout, J. C., Gabery, S., Churchyard, A., Chua, P., Egan, G. F., . . . Poudel, G. R. (2016). Subjective sleep problems in Huntington's disease: a pilot investigation of the relationship to brain structure, neurocognitive, and neuropsychiatric function. *Journal of the Neurological Sciences, 364*, 148-153.

Bamford, K. A., Caine, E. D., Kido, D. K., Cox, C., & Shoulson, I. (1995). A prospective evaluation of cognitive decline in early Huntington's disease. *Functional and radiographic correlates, 45*(10), 1867-1873. doi:10.1212/wnl.45.10.1867

Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J.-P. (2018). Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology, 43*(8), 1660-1666.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology(). *Archives of Clinical Neuropsychology, 27*(3), 362-373. doi:10.1093/arclin/acs027

Beebe, D. W. (2011). Cognitive, behavioral, and functional consequences of inadequate sleep in children and adolescents. *Pediatric Clinics, 58*(3), 649-665.

Begeti, F., Schwab, L. C., Mason, S. L., & Barker, R. A. (2016). Hippocampal dysfunction defines disease onset in Huntington's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 87*(9), 975-981.

Beglinger, L. J., O'Rourke, J. J. F., Wang, C., Langbehn, D. R., Duff, K., & Paulsen, J. S. (2010). Earliest functional declines in Huntington disease. *Psychiatry Research, 178*(2), 414-418. doi:10.1016/j.psychres.2010.04.030

Beglinger, L. J., Prest, L., Mills, J. A., Paulsen, J. S., Smith, M. M., Gonzalez-Alegre, P., . . . Uc, E. Y. (2012). Clinical predictors of driving status in Huntington's disease. *Movement Disorders, 27*(9), 1146-1152.

Bei, B., Wiley, J. F., Trinder, J., & Manber, R. (2016). Beyond the mean: a systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. *Sleep medicine reviews, 28*, 108-124.

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal, 38*(3), 218-226. doi:10.1037/prj0000130

Beste, C., Stock, A. K., Ness, V., Hoffmann, R., Lukas, C., & Saft, C. (2013). A novel cognitive-neurophysiological state biomarker in premanifest Huntington's disease validated on longitudinal data. *Scientific reports, 3*. doi:10.1038/srep01797

Blumenstock, S., & Dudanova, I. (2020). Cortical and Striatal Circuits in Huntington's Disease. *Frontiers in Neuroscience, 14*, 82.

Boe, A. J., McGee Koch, L. L., O'Brien, M. K., Shawen, N., Rogers, J. A., Lieber, R. L., . . . Jayaraman, A. (2019). Automating sleep stage classification using wireless, wearable sensors. *npj Digital Medicine, 2*(1), 131. doi:10.1038/s41746-019-0210-1

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences, 33*(1), 10-16.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, research & practice, 16*(3), 252.

Born, J., & Wagner, U. (2004). Awareness in memory: being explicit about the role of sleep. *Trends in cognitive sciences, 8*(6), 242-244.

Boulos, M. N. K., Wheeler, S., Tavares, C., & Jones, R. (2011). How smartphones are changing the face of mobile and participatory healthcare: An overview, with example from eCAALYX. *Biomedical Engineering Online, 10*. doi:10.1186/1475-925X-10-24

Brandt, J., & Benedict, R. H. (2001). *Hopkins verbal learning test, revised: professional manual*: Psychological Assessment Resources.

Brosnan, M. B., Sabaroedin, K., Silk, T., Genc, S., Newman, D. P., Loughnane, G. M., . . . Bellgrove, M. A. (2020). Evidence accumulation during perceptual decisions in humans varies as a function of dorsal frontoparietal organization. *Nature human behaviour*, 1-12.

Brown, B., Peiffer, J., & Martins, R. (2013). Multiple effects of physical activity on molecular and cognitive signs of brain aging: can exercise slow neurodegeneration and delay Alzheimer's disease? *Molecular Psychiatry*, *18*(8), 864-874.

Buffalo, E. A., Bellgowan, P. S., & Martin, A. (2006). Distinct roles for medial temporal lobe structures in memory for objects and their locations. *Learning & Memory*, *13*(5), 638-643.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*: Sage.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253.

Busse, M., Quinn, L., Debono, K., Jones, K., Collett, J., Playle, R., . . . the Members of the, C.-H. D. M. G. (2013). A Randomized Feasibility Study of a 12-Week Community-Based Exercise Program for People With Huntington's Disease. *pre, 37*(4). Retrieved from https://journals.lww.com/jnpt/Fulltext/2013/12000/A_Randomized_Feasibility_Study_of_a_12_Week.2.aspx

Campodonico, J. R., Aylward, E., Codori, A.-M., Young, C., Krafft, L., Magdalinski, M., . . . Brandt, J. (1998). When does Huntington's disease begin? *Journal of the International Neuropsychological Society, 4*(5), 467-473.

Cardoso, F. (2009). Huntington Disease and Other Choreas. *Neurologic Clinics, 27*(3), 719-736. doi:10.1016/j.ncl.2009.04.001

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data. *JAMA, 313*(6), 625-626. doi:10.1001/jama.2014.17841

Chamberlain, S. R., Robbins, T. W., Winder-Rhodes, S., Müller, U., Sahakian, B. J., Blackwell, A. D., & Barnett, J. H. (2011). Translational approaches to frontostriatal dysfunction in attention-deficit/hyperactivity disorder using a computerized neuropsychological battery. *Biological Psychiatry, 69*(12), 1192-1203.

Chesson Jr, M., Coleman, M., Lee-Chiong, M., & Pancer, D. (2007). Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep, 30*(4), 519.

Choi-Kwon, S., & Jeon, B. (2017). Are there any differences in subjective and objective sleep measurement? *Sleep Medicine, 40*, e65. doi:https://doi.org/10.1016/j.sleep.2017.11.184

Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults a meta-analytic study. *Psychological science, 14*(2), 125-130.

Collerton, J., Collerton, D., Arai, Y., Barrass, K., Eccles, M., Jagger, C., . . . Von Zglinicki, T. (2007). A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ pilot study. *Journal of the American Geriatrics Society, 55*(10), 1630-1635. doi:10.1111/j.1532-5415.2007.01379.x

Collie, A., Makdissi, M., Maruff, P., Bennell, K., & McCrory, P. (2006). Cognition in the days following concussion: comparison of symptomatic versus asymptomatic athletes. *Journal of Neurology, Neurosurgery & Psychiatry, 77*(2), 241-245.

Coulter, A., & Ellins, J. (2006). *Patient-focused interventions: a review of the evidence* (Vol. 1): Health Foundation London.

Cruickshank, T. M., Thompson, J. A., Domínguez D, J. F., Reyes, A. P., Bynevelt, M., Georgiou-Karistianis, N., . . . Ziman, M. R. (2015). The effect of multidisciplinary rehabilitation on brain structure and cognition in Huntington's disease: an exploratory study. *Brain and Behavior, 5*(2), n/a-n/a. doi:10.1002/brb3.312

Cuturic, M., Abramson, R. K., Vallini, D., Frank, E. M., & Shamsnia, M. (2009). Sleep Patterns in Patients With Huntington's Disease and Their Unaffected First-Degree Relatives: A Brief Report. *Behavioral Sleep Medicine, 7*(4), 245-254. doi:10.1080/15402000903190215

Daneshi, A., Azarnoush, H., & Towhidkhah, F. (2020). A one-boundary drift-diffusion model for time to collision estimation in a simple driving task. *Journal of Cognitive Psychology, 32*(1), 67-81.

Daneshi, A., Azarnoush, H., Towhidkhah, F., Gohari, A., & Ghazizadeh, A. (2019). Drift-diffusion explains response variability and capacity for tracking objects. *Scientific reports, 9*(1), 1-15.

Daniëls, N. E. M., Bartels, S. L., Verhagen, S. J. W., Van Knippenberg, R. J. M., De Vugt, M. E., & Delespaul, P. A. E. G. (2020). Digital assessment of working memory and processing speed in everyday life: Feasibility, validation, and lessons-learned. *Internet Interventions, 19*, 100300. doi:https://doi.org/10.1016/j.invent.2019.100300

Darby, D. G., Fredrickson, J., Pietrzak, R. H., Maruff, P., Woodward, M., & Brodtmann, A. (2014). Reliability and usability of an internet-based computerized cognitive testing

battery in community-dwelling older people. *Computers in Human Behavior, 30,* 199-205. doi:10.1016/j.chb.2013.08.009

de Zambotti, M., Claudatos, S., Inkelis, S., Colrain, I. M., & Baker, F. C. (2015). Evaluation of a consumer fitness-tracking device to assess sleep in adults. *Chronobiology International, 32*(7), 1024-1028.

Devos, H., Nieuwboer, A., Tant, M., De Weerdt, W., & Vandenberghe, W. (2012). Determinants of fitness to drive in Huntington disease. *Neurology, 79*(19), 1975-1982.

Diago, E. B., Martínez-Horta, S., Lasaosa, S. S., Alebesque, A. V., Pérez-Pérez, J., Kulisevsky, J., & del Val, J. L. (2018). Circadian rhythm, cognition, and mood disorders in Huntington's disease. *Journal of Huntington's disease, 7*(2), 193-198.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews. Neuroscience, 11*(2), 114-126. Retrieved from http://dx.doi.org/10.1038/nrn2762

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., . . . Grainger, J. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PLoS ONE, 6*(9). doi:10.1371/journal.pone.0024974

Duff, K., Paulsen, J. S., Beglinger, L. J., Langbehn, D. R., Wang, C., Stout, J. C., . . . Queller, S. (2010). "Frontal" behaviors before the diagnosis of Huntington's disease and their relationship to markers of disease progression: Evidence of early lack of awareness. *Journal of Neuropsychiatry and Clinical Neurosciences, 22*(2), 196-207. doi:10.1176/appi.neuropsych.22.2.196

Eddy, C. M., & Rickards, H. E. (2013). Impact of cognitive and behavioural changes on quality of life in Huntington's disease. *Basal Ganglia, 3*(2), 123-126. doi:10.1016/j.baga.2013.01.085

Eddy, C. M., & Rickards, H. E. (2015). Cognitive deficits predict poorer functional capacity in huntington's disease: But what is being measured? *Neuropsychology, 29*(2), 268-273. doi:10.1037/neu0000134

Ellenbogen, J. M. (2005). Cognitive benefits of sleep and their loss due to sleep deprivation. *Neurology, 64*(7), E25-E27.

Esteban-Cornejo, I., Tejero-Gonzalez, C. M., Sallis, J. F., & Veiga, O. L. (2015). Physical activity and cognition in adolescents: A systematic review. *Journal of Science and Medicine in Sport, 18*(5), 534-539. doi:https://doi.org/10.1016/j.jsams.2014.07.007

Estudillo-Valderrama, M. Á., Roa, L. M., Reina-Tosina, J., & Naranjo-Hernández, D. (2009). Design and implementation of a distributed fall detection system-personal server. *IEEE Transactions on Information Technology in Biomedicine, 13*(6), 874-881. doi:10.1109/TITB.2009.2031316

Eve, C. T., & De Jager, C. A. (2014). Piloting and validation of a novel self-administered online cognitive screening tool in normal older persons: The Cognitive Function Test. *International Journal of Geriatric Psychiatry, 29*(2), 198-206. doi:10.1002/gps.3993

Farrell, H. J., Andrews, S. C., Ryan, N. P., Davis, M.-C., Gordon, S., Stout, J. C., & Fisher, F. (2019). Cognitive Fitness to Drive in Huntington's Disease: Assessing the Clinical Utility of DriveSafe DriveAware. *Journal of Huntington's disease, 8*(1), 87-95.

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 42.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Frank, M. J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks, 19*(8), 1120-1136.

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science, 318*(5854), 1309-1312.

Fredrickson, J., Maruff, P., Woodward, M., Moore, L., Fredrickson, A., Sach, J., & Darby, D. (2010). Evaluation of the Usability of a Brief Computerized Cognitive Screening Test in Older People for Epidemiological Studies. *Neuroepidemiology, 34*(2), 65-75. doi:http://dx.doi.org/10.1159/000264823

Fritz, N. E., Rao, A. K., Kegelmeyer, D., Kloos, A., Busse, M., Hartel, L., . . . Quinn, L. (2017). Physical Therapy and Exercise Interventions in Huntington's Disease: A Mixed Methods Systematic Review. *Journal of Huntington's disease, 6*, 217-235. doi:10.3233/JHD-170260

Ghilardi, M., Silvestri, G., Feigin, A., Mattis, P., Zgaljardic, D., Moisello, C., . . . Eidelberg, D. (2008). Implicit and explicit aspects of sequence learning in pre-symptomatic Huntington's disease. *Parkinsonism & related disorders, 14*(6), 457-464.

Glikmann-Johnston, Y., Carmichael, A. M., Mercieca, E.-C., & Stout, J. C. (2019). 'Real-life'hippocampal-dependent spatial memory impairments in Huntington's disease. *Cortex, 119*, 46-60.

Glikmann-Johnston, Y., Fink, K. D., Deng, P., Torrest, A., & Stout, J. C. (2019). Spatial memory in Huntington's disease: A comparative review of human and animal data. *Neuroscience & Biobehavioral Reviews, 98*, 194-207. doi:https://doi.org/10.1016/j.neubiorev.2019.01.015

Glikmann-Johnston, Y., Halag-Milo, T., Hendler, T., & Gazit, T. (2020). Functional MRI (fMRI) human mapping for indicating hippocampal content-specific responsivity and laterality. *Neuropsychology*.

Goodman, A. O., & Barker, R. A. (2010). How vital is sleep in Huntington's disease? *Journal of Neurology, 257*(6), 882-897.

Goodman, A. O., Morton, A. J., & Barker, R. A. (2010). Identifying sleep disturbances in Huntington's disease using a simple disease-focused questionnaire. *PLoS currents, 2*.

Goodman, A. O., Rogers, L., Pilsworth, S., McAllister, C. J., Shneerson, J. M., Morton, A. J., & Barker, R. A. (2011). Asymptomatic sleep abnormalities are a common early feature in patients with Huntington's disease. *Current Neurology and Neuroscience Reports, 11*(2), 211-217.

Hagelkruys, D., Böhm, C., Motschnig, R., Kertzman, S., Sirota, A., Lerer, B., & Gyarmathy, É. (2016). Adapting a stand-alone computerized cognitive test battery for online use–A case-study in the context of users with special needs. *Computers in Human Behavior, 63*, 757-768.

Hagströmer, M., Oja, P., & Sjöström, M. (2006). The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutrition, 9*(06), 755-762.

Hamer, M., & Chida, Y. (2009). Physical activity and risk of neurodegenerative disease: a systematic review of prospective evidence. *Psychological Medicine, 39*(1), 3-11. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-62249177925&partnerID=40&md5=eb3dc1d24ece82f1ed5081b7faac9c83

Hamo, N., Abramovitch, A., & Zohar, A. (2018). A computerized neuropsychological evaluation of cognitive functions in a subclinical obsessive-compulsive sample. *Journal of Behavior Therapy and Experimental Psychiatry*.

Hansotia, P., Wall, R., & Berendes, J. (1985). Sleep disturbances and severity of Huntington's disease. *Neurology, 35*(11), 1672-1672.

Happe, S., & Trenkwalder, C. (2002). Movement disorders in sleep: Gilles de la Tourette syndrome, Huntington's disease, and dystonia. *Somnologie-Schlafforschung und Schlafmedizin, 6*(2), 63-67.

Hennig, B. L., Kaplan, R. F., Nowicki, A. E., Barclay, J. E., & Gertsberg, A. G. (2014). We can predict when driving is no longer safe for people who have HD using standard neuropsychological measures. *Journal of Huntington's disease*, *3*(4), 351-353.

Henson, P., Peck, P., & Torous, J. (2019). Considering the therapeutic alliance in digital mental health interventions. *Harvard Review of Psychiatry, 27*(4), 268-273.

Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., & Torous, J. (2019). Digital mental health apps and the therapeutic alliance: initial review. *BJPsych Open, 5*(1).

Herz, D. M., Little, S., Pedrosa, D. J., Tinkhauser, G., Cheeran, B., Foltynie, T., . . . Brown, P. (2018). Mechanisms Underlying Decision-Making as Revealed by Deep-Brain Stimulation in Patients with Parkinson's Disease. *Current Biology, 28*(8), 1169-1178.e1166. doi:https://doi.org/10.1016/j.cub.2018.02.057

Herz, Damian M., Zavala, Baltazar A., Bogacz, R., & Brown, P. (2016). Neural Correlates of Decision Thresholds in the Human Subthalamic Nucleus. *Current Biology, 26*(7), 916-920. doi:https://doi.org/10.1016/j.cub.2016.01.051

Herzog-Krzywoszanska, R., & Krzywoszanski, L. (2019). Sleep disorders in Huntington's disease. *Frontiers in psychiatry, 10*, 221.

Heyn, P., Abreu, B. C., & Ottenbacher, K. J. (2004). The effects of exercise training on elderly persons with cognitive impairment and dementia: a meta-analysis. *Archives of Physical Medicine and Rehabilitation, 85*(10), 1694-1704.

Hirsch, M. A., Iyer, S. S., & Sanjak, M. (2016). Exercise-induced neuroplasticity in human Parkinson's disease: What is the evidence telling us? *Parkinsonism & related disorders, 22*, S78-S81. doi:https://doi.org/10.1016/j.parkreldis.2015.09.030

Ho, A. K., Robbins, A. O., & Barker, R. A. (2006). Huntington's disease patients have selective problems with insight. *Movement Disorders, 21*(3), 385-389.

Ho, A. K., Sahakian, B. J., Brown, R. G., Barker, R. A., Hodges, J. R., Ané, M. N., . . . Bodner, T. (2003). Profile of cognitive progression in early Huntington's disease. *Neurology, 61*(12), 1702-1706. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-9144259577&partnerID=40&md5=bcef279b278aa75ae429fb2ef0fbd4c2

Hoth, K. F., Paulsen, J. S., Moser, D. J., Tranel, D., Clark, L. A., & Bechara, A. (2007). Patients with Huntington's disease have impaired awareness of cognitive, emotional, and functional abilities. *Journal of Clinical and Experimental Neuropsychology, 29*(4), 365-376. doi:10.1080/13803390600718958

Hötting, K., & Röder, B. (2013). Beneficial effects of physical exercise on neuroplasticity and cognition. *Neuroscience & Biobehavioral Reviews, 37*(9), 2243-2257.

Huang, Y.-T., Georgiev, D., Foltynie, T., Limousin, P., Speekenbrink, M., & Jahanshahi, M. (2015). Different effects of dopaminergic medication on perceptual decision-making in Parkinson's disease as a function of task difficulty and speed–accuracy instructions. *Neuropsychologia, 75*, 577-587.

Huntington Study Group. (1996). Unified Huntington's disease rating scale: reliability and consistency.

Iverson, G. L., Brooks, B. L., Langenecker, S. A., & Young, A. H. (2011). Identifying a cognitive impairment subgroup in adults with mood disorders. *Journal of Affective Disorders, 132*(3), 360-367.

Jacobs, M., Hart, E. P., & Roos, R. A. (2017). Driving with a neurodegenerative disorder: an overview of the current literature. *Journal of Neurology, 264*(8), 1678-1696.

Jewett, M. E., Dijk, D.-J., Kronauer, R. E., & Dinges, D. F. (1999). Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep: Journal of Sleep Research & Sleep Medicine*.

Johnson, P. L., Potts, G. F., Sanchez-Ramos, J., & Cimino, C. R. (2017). Self-reported impulsivity in Huntington's disease patients and relationship to executive dysfunction and reward responsiveness. *Journal of Clinical and Experimental Neuropsychology, 39*(7), 694-706. doi:10.1080/13803395.2016.1257702

Jongstra, S., Wijsman, L. W., Cachucho, R., Hoevenaar-Blom, M. P., Mooijaart, S. P., & Richard, E. (2017a). Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. *JMIR mHealth and uHealth, 5*(5), e68.

Jongstra, S., Wijsman, L. W., Cachucho, R., Hoevenaar-Blom, M. P., Mooijaart, S. P., & Richard, E. (2017b). Cognitive Testing in People at Increased Risk of Dementia Using a Smartphone App: The iVitality Proof-of-Principle Study. *JMIR Mhealth Uhealth, 5*(5), e68. doi:10.2196/mhealth.6939

Kahn-Greene, E. T., Lipizzi, E. L., Conrad, A. K., Kamimori, G. H., & Killgore, W. D. S. (2006). Sleep deprivation adversely affects interpersonal responses to frustration. *Personality and Individual Differences, 41*(8), 1433-1443.

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology, 117*(7), 1574-1581.

Kang, S., Kim, Y., Byun, W., Suk, J., & Lee, J.-M. (2019). Comparison of a Wearable Tracker with Actigraph for Classifying Physical Activity Intensity and Heart Rate in

Children. *International Journal of Environmental Research and Public Health, 16*(15), 2663.

Kaye, J. A., Maxwell, S. A., Mattek, N., Hayes, T. L., Dodge, H., Pavel, M., . . . Zitzelberger, T. A. (2011). Intelligent Systems for Assessing Aging Changes: Home-Based, Unobtrusive, and Continuous Assessment of Aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66B*(suppl 1), i180-i190. doi:10.1093/geronb/gbq095

Killgore, W. D. S. (2010). Effects of sleep deprivation on cognition. *Human Sleep and Cognition: Basic Research, 185*, 105.

Killgore, W. D. S., Balkin, T. J., & Wesensten, N. J. (2006). Impaired decision making following 49 h of sleep deprivation. *Journal of Sleep Research, 15*(1), 7-13.

Klimova, B. (2017). Mobile Phone Apps in the Management and Assessment of Mild Cognitive Impairment and/or Mild-to-Moderate Dementia: An Opinion Article on Recent Findings. *Frontiers in Human Neuroscience, 11*(461). doi:10.3389/fnhum.2017.00461

Koo, B. M., & Vizer, L. M. (2019). Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review. *Innovation in Aging, 3*(1). doi:10.1093/geroni/igy038

Kramer, A. F., Hahn, S., Cohen, N. J., Banich, M. T., McAuley, E., Harrison, C. R., . . . Colcombe, A. (1999). Ageing, fitness and neurocognitive function. *Nature, 400*(6743), 418-419. Retrieved from http://dx.doi.org/10.1038/22682

Landry, G., Best, J., & Liu-Ambrose, T. (2015). Measuring sleep quality in older adults: a comparison using subjective and objective methods. *Frontiers in Aging Neuroscience, 7*(166). doi:10.3389/fnagi.2015.00166

Lang, C., Brand, S., Feldmeth, A. K., Holsboer-Trachsler, E., Pühse, U., & Gerber, M. (2013). Increased self-reported and objectively assessed physical activity predict sleep

quality among adolescents. *Physiology & Behavior, 120*, 46-53. doi:https://doi.org/10.1016/j.physbeh.2013.07.001

Lautenschlager, N. T., Cox, K., & Kurz, A. F. (2010). Physical activity and mild cognitive impairment and Alzheimer's disease. *Current Neurology and Neuroscience Reports, 10*(5), 352-358. doi:10.1007/s11910-010-0121-7

Lazar, A. S., Panin, F., Goodman, A. O., Lazic, S. E., Lazar, Z. I., Mason, S. L., . . . Singh, P. (2015). Sleep deficits but no metabolic deficits in premanifest H untington's disease. *Annals of Neurology, 78*(4), 630-648.

Leidy, N. K., & Vernon, M. (2008). Perspectives on patient-reported outcomes. *Pharmacoeconomics, 26*(5), 363-370.

Lezak, M. D., Howieson, D., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York: Oxford University Press.

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences, 1129*(1), 305-322.

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin, 136*(3), 375.

Lim, Y. Y., Ellis, K. A., Harrington, K., Ames, D., Martins, R. N., Masters, C. L., . . . Maruff, P. (2012). Use of the CogState brief battery in the assessment of Alzheimer's disease related cognitive impairment in the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. *Journal of Clinical and Experimental Neuropsychology, 34*(4), 345-358. doi:10.1080/13803395.2011.643227

Lipsmeier, F., Taylor, K. I., Kilchenmann, T., Wolf, D., Scotland, A., Schjodt-Eriksen, J., . . . Jin, L. (2018). Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Movement Disorders, 33*(8), 1287-1297.

Long, J. D., Lee, J.-M., Aylward, E. H., Gillis, T., Mysore, J. S., Elneel, K. A., . . . Gusella, J. F. (2018). Genetic modification of Huntington disease acts early in the prediagnosis phase. *The American Journal of Human Genetics, 103*(3), 349-357.

Makdissi, M., Collie, A., Maruff, P., Darby, D. G., Bush, A., McCrory, P., & Bennell, K. (2001). Computerised cognitive assessment of concussed Australian Rules footballers. *British Journal of Sports Medicine, 35*(5), 354-360. doi:10.1136/bjsm.35.5.354

Marinus, J., Visser, M., van Hilten, J. J., Lammers, G. J., & Stiggelbout, A. M. (2003). Assessment of Sleep and Sleepiness in Parkinson Disease. *Sleep, 26*(8), 1049-1054. doi:10.1093/sleep/26.8.1049

Maroof, D. A., Gross, A. L., & Brandt, J. (2011). Modeling longitudinal change in motor and cognitive processing speed in presymptomatic Huntington's disease. *Journal of Clinical and Experimental Neuropsychology, 33*(8), 901-909. doi:10.1080/13803395.2011.574606

Maskevich, S., Jumabhoy, R., Dao, P. D., Stout, J. C., & Drummond, S. (2017). Pilot Validation of Ambulatory Activity Monitors for Sleep Measurement in Huntington's Disease Gene Carriers. *Journal of Huntington's disease, 6*(3), 249-253.

McCrae, C. S., Vatthauer, K. E., Dzierzewski, J. M., & Marsiske, M. (2012). Habitual sleep, reasoning, and processing speed in older adults with sleep complaints. *Cognitive therapy and research, 36*(2), 156-164.

Mo, C., Hannan, A. J., & Renoir, T. (2015). Environmental factors as modulators of neurodegeneration: Insights from gene–environment interactions in Huntington's disease. *Neuroscience & Biobehavioral Reviews, 52*, 178-192. doi:https://doi.org/10.1016/j.neubiorev.2015.03.003

Mohr, E., Walker, D., Randolph, C., Sampson, M., & Mendis, T. (1996). Utility of clinical trial batteries in the measurement of Alzheimer's and Huntington's dementia. *International Psychogeriatrics, 8*(03), 397-411.

Montgomery-Downs, H. E., Insana, S. P., & Bond, J. A. (2012). Movement toward a novel activity monitoring device. *Sleep and Breathing, 16*(3), 913-917.

Moore, R. C., Swendsen, J., & Depp, C. A. (2017). Applications for self-administered mobile cognitive assessments in clinical research: A systematic review. *International Journal of Methods in Psychiatric Research, 26*(4), e1562. doi:10.1002/mpr.1562

Mörkl, S., Müller, N. J., Blesl, C., Wilkinson, L., Tmava, A., Wurm, W., . . . Painold, A. (2016). Problem solving, impulse control and planning in patients with early- and late-stage Huntington's disease. *European Archives of Psychiatry and Clinical Neuroscience, 266*(7), 663-671. doi:10.1007/s00406-016-0707-4

Morton, A. J. (2013). Circadian and sleep disorder in Huntington's disease. *Experimental Neurology, 243*, 34-44. doi:10.1016/j.expneurol.2012.10.014

Morton, A. J., Wood, N. I., Hastings, M. H., Hurelbrink, C., Barker, R. A., & Maywood, E. S. (2005). Disintegration of the sleep-wake cycle and circadian timing in Huntington's disease. *The Journal of neuroscience, 25*(1), 157-163.

Moss, D. J. H., Pardiñas, A. F., Langbehn, D., Lo, K., Leavitt, B. R., Roos, R., . . . Santos, R. D. (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *The Lancet Neurology, 16*(9), 701-711.

Mulder, M., Van Maanen, L., & Forstmann, B. (2014). Perceptual decision neurosciences–a model-based review. *Neuroscience, 277*, 872-884.

Nithianantharajah, J., Barkus, C., Murphy, M., & Hannan, A. J. (2008). Gene–environment interactions modulating cognitive function and molecular correlates of synaptic

plasticity in Huntington's disease transgenic mice. *Neurobiology of Disease, 29*(3), 490-504.

O'Callaghan, C., Bertoux, M., & Hornberger, M. (2014). Beyond and below the cortex: the contribution of striatal dysfunction to cognition and behaviour in neurodegeneration. *Journal of Neurology, Neurosurgery & Psychiatry*, *85*(4), 371-378.

Ocloo, J., & Matthews, R. (2016). From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Quality & Safety, 25*(8), 626. doi:10.1136/bmjqs-2015-004839

Paillard, T., Rolland, Y., & de Souto Barreto, P. (2015). Protective effects of physical exercise in Alzheimer's disease and Parkinson's disease: a narrative review. *Journal of clinical neurology, 11*(3), 212-219.

Pallier, P. N., Maywood, E. S., Zheng, Z., Chesham, J. E., Inyushkin, A. N., Dyball, R., . . . Morton, A. J. (2007). Pharmacological imposition of sleep slows cognitive decline and reverses dysregulation of circadian gene expression in a transgenic mouse model of Huntington's disease. *The Journal of neuroscience, 27*(29), 7869-7878.

Pallier, P. N., & Morton, A. J. (2009). Management of sleep/wake cycles improves cognitive function in a transgenic mouse model of Huntington's disease. *Brain Research, 1279*, 90-98. doi:http://dx.doi.org/10.1016/j.brainres.2009.03.072

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of technology in neuropsychological assessment. *The Clinical Neuropsychologist, 27*(8), 1328-1361.

Paulsen, J. S., Langbehn, D. R., Stout, J. C., Aylward, E., Ross, C. A., Nance, M., . . . Hayden, M. (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *Journal of Neurology, Neurosurgery & Psychiatry, 79*(8), 874-880. doi:10.1136/jnnp.2007.128728

Paulsen, J. S., Long, J. D., Johnson, H. J., Aylward, E. H., Ross, C. A., Williams, J. K., . . . Harrington, D. L. (2014). Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: a decade of the PREDICT-HD study. *Frontiers in Aging Neuroscience, 6*, 78.

Paulsen, J. S., Smith, M. M., & Long, J. D. (2013). Cognitive decline in prodromal Huntington disease: Implications for clinical trials. *Journal of Neurology, Neurosurgery and Psychiatry, 84*(11), 1233-1239. doi:10.1136/jnnp-2013-305114

Payne, J. M., D'Antoine, H. A., France, K. E., McKenzie, A. E., Henley, N., Bartu, A. E., . . . Bower, C. (2011). Collaborating with consumer and community representatives in health and medical research in Australia: results from an evaluation. *Health Research Policy and Systems, 9*(1), 18. doi:10.1186/1478-4505-9-18

Peinemann, A., Schuller, S., Pohl, C., Jahn, T., Weindl, A., & Kassubek, J. (2005). Executive dysfunction in early stages of Huntington's disease is associated with striatal and insular atrophy: a neuropsychological and voxel-based morphometric study. *Journal of the Neurological Sciences, 239*(1), 11-19.

Penney Jr, J. B., Vonsattel, J. P., Macdonald, M. E., Gusella, J. F., & Myers, R. H. (1997). CAG repeat number governs the development rate of pathology in Huntington's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 41*(5), 689-692.

Possin, K. L., Kim, H., Geschwind, M. D., Moskowitz, T., Johnson, E. T., Sharon, J. S., . . . Finkbeiner, S. (2017). Egocentric and allocentric visuospatial working memory in premotor Huntington's disease: A double dissociation with caudate and hippocampal volumes. *Neuropsychologia, 101*, 57-64.

Poudel, G. R., Stout, J. C., Domínguez D, J. F., Salmon, L., Churchyard, A., Chua, P., . . . Egan, G. F. (2014). White matter connectivity reflects clinical and cognitive status in

Huntington's disease. *Neurobiology of Disease, 65*, 180-187. doi:https://doi.org/10.1016/j.nbd.2014.01.013

Pringsheim, T., Wiltshire, K., Day, L., Dykeman, J., Steeves, T., & Jette, N. (2012). The incidence and prevalence of Huntington's disease: A systematic review and meta-analysis. *Movement Disorders, 27*(9), 1083-1091. doi:10.1002/mds.25075

Quinn, L., Debono, K., Dawes, H., Rosser, A. E., Nemeth, A. H., Rickards, H., . . . Kelson, M. J. (2014). Task-specific training in Huntington disease: a randomized controlled feasibility trial. *Physical Therapy, 94*(11), 1555-1568.

Quinn, L., Hamana, K., Kelson, M., Dawes, H., Collett, J., Townson, J., . . . Busse, M. (2016). A randomized, controlled trial of a multi-modal exercise intervention in Huntington's disease. *Parkinsonism & related disorders, 31*, 46-52. doi:https://doi.org/10.1016/j.parkreldis.2016.06.023

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873-922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences, 20*(4), 260-281.

Rentz, D. M., Dekhtyar, M., Sherman, J., Burnham, S., Blacker, D., Aghjayan, S. L., . . . Chenhall, T. (2016). The feasibility of at-home iPad cognitive testing for use in clinical trials. *The journal of prevention of Alzheimer's disease, 3*(1), 8.

Riemann, D., Berger, M., & Voderholzer, U. (2001). Sleep and depression—results from psychobiological studies: an overview. *Biological Psychology, 57*(1-3), 67-103.

Rodrigues, F. B., & Wild, E. J. (2020). Huntington's Disease Clinical Trials Corner: April 2020. *J Huntingtons Dis*. doi:10.3233/jhd-200002

Rosas, H., Hevelone, N., Zaleta, A., Greve, D., Salat, D., & Fischl, B. (2005). Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. *Neurology, 65*(5), 745-747.

Ross, C. A., Aylward, E. H., Wild, E. J., Langbehn, D. R., Long, J. D., Warner, J. H., . . . Tabrizi, S. J. (2014). Huntington disease: Natural history, biomarkers and prospects for therapeutics. *Nature Reviews Neurology, 10*(4), 204-216. doi:10.1038/nrneurol.2014.24

Ross, C. A., Pantelyat, A., Kogan, J., & Brandt, J. (2014). Determinants of functional disability in Huntington's disease: role of cognitive and motor dysfunction. *Movement Disorders, 29*(11), 1351-1358.

Rovio, S., Kåreholt, I., Helkala, E.-L., Viitanen, M., Winblad, B., Tuomilehto, J., . . . Kivipelto, M. (2005). Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *The Lancet Neurology, 4*(11), 705-711. doi:http://dx.doi.org/10.1016/S1474-4422(05)70198-8

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. *Sleep medicine reviews, 15*(4), 259-267. doi:http://dx.doi.org/10.1016/j.smrv.2010.10.001

Schlegel, R. E., & Gilliland, K. (2007). Development and quality assurance of computer-based assessment batteries. *Archives of Clinical Neuropsychology, 22*(SUPPL. 1), 49-61. doi:10.1016/j.acn.2006.10.005

Shambroom, J. R., Fábregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research, 21*(2), 221-230.

Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: a meta-analysis. *Pediatric exercise science, 15*(3), 243-256.

Silvestri, R., Raffaele, M., De Domenico, P., Tisano, A., Mento, G., Casella, C., . . . Di Perri, R. (1995). Sleep features in Tourette's syndrome, neuroacanthocytosis and Huntington's chorea. *Neurophysiologie Clinique/Clinical Neurophysiology, 25*(2), 66-77.

Song, S.-K., Yoshino, J., Le, T. Q., Lin, S.-J., Sun, S.-W., Cross, A. H., & Armstrong, R. C. (2005). Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage, 26*(1), 132-140.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 7*(3), 280-292. doi:10.1016/j.jalz.2011.03.003

Starkstein, S. E., Brandt, J., Bylsma, F., Peyser, C., Folstein, M., & Folstein, S. E. (1992). Neuropsychological correlates of brain atrophy in Huntington's disease: a magnetic resonance imaging study. *Neuroradiology, 34*(6), 487-489. doi:10.1007/BF00598956

Stout, J. C., Andrews, S. C., & Glikmann-Johnston, Y. (2017). Cognitive assessment in Huntington disease clinical drug trials. In *Handbook of Clinical Neurology* (Vol. 144, pp. 227-244): Elsevier.

Stout, J. C., Jones, R., Labuschagne, I., O'Regan, A. M., Say, M. J., Dumas, E. M., . . . Frost, C. (2012). Evaluation of longitudinal 12 and 24 month cognitive outcomes in premanifest and early Huntington's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 83*(7), 687-694. doi:10.1136/jnnp-2011-301940

Stout, J. C., Queller, S., Baker, K. N., Cowlishaw, S., Sampaio, C., Fitzer-Attas, C., . . . Thompson, J. (2014). HD-CAB: A cognitive assessment battery for clinical trials in

Huntington's disease. *Movement Disorders, 29*(10), 1281-1288. doi:10.1002/mds.25964

Stout, J. C., Rodawalt, W. C., & Siemers, E. R. (2001). Risky decision making in Huntington's disease. *Journal of the International Neuropsychological Society, 7*(1), 92-101.

Sweeney, J. A., Kmiec, J. A., & Kupfer, D. J. (2000). Neuropsychologic impairments in bipolar and unipolar mood disorders on the CANTAB neurocognitive battery. *Biological Psychiatry, 48*(7), 674-684.

Tabrizi, S. J., Reilmann, R., Roos, R. A. C., Durr, A., Leavitt, B., Owen, G., . . . Langbehn, D. R. (2012). Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *The Lancet Neurology, 11*(1), 42-53. doi:http://dx.doi.org/10.1016/S1474-4422(11)70263-0

Tabrizi, S. J., Scahill, R. I., Owen, G., Durr, A., Leavitt, B. R., Roos, R. A., . . . Langbehn, D. R. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: Analysis of 36-month observational data. *The Lancet Neurology, 12*(7), 637-649. doi:10.1016/S1474-4422(13)70088-7

Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport, 17*(5), 496-500. doi:10.1016/j.jsams.2013.10.241

Taylor, N., & Bramble, D. (1997). Sleep disturbance and Huntingdon's disease. *The British Journal of Psychiatry, 171*(4), 393-393.

Thompson, J. A., Cruickshank, T., Penailillo, L., Lee, J., Newton, R., Barker, R., & Ziman, M. R. (2013). The effects of multidisciplinary rehabilitation in patients with early-to-

middle-stage Huntington's disease: a pilot study. *European Journal of Neurology, 20*(9), 1325-1329.

Thomson, E. A., Nuss, K., Comstock, A., Reinwald, S., Blake, S., Pimentel, R. E., . . . Li, K. (2019). Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *Journal of Sports Sciences, 37*(12), 1411-1419.

Tierney, M. C., Naglie, G., Upshur, R., Moineddin, R., Charles, J., & Liisa Jaakkimainen, R. (2014). Feasibility and validity of the self-administered computerized assessment of mild cognitive impairment with older primary care patients. *Alzheimer Disease and Associated Disorders, 28*(4), 311-319. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84914164758&partnerID=40&md5=ee40c979ae78d87b6f471a65b12aad14

Toh, E. A., MacAskill, M. R., Dalrymple-Alford, J. C., Myall, D. J., Livingston, L., Macleod, S. A. D., & Anderson, T. J. (2014). Comparison of cognitive and UHDRS measures in monitoring disease progression in Huntington's disease: A 12-month longitudinal study. *Translational Neurodegeneration, 3*(1). doi:10.1186/2047-9158-3-15

Tonetti, L., Pasquini, F., Fabbri, M., Belluzzi, M., & Natale, V. (2008). Comparison of two different actigraphs with polysomnography in healthy young subjects. *Chronobiology International, 25*(1), 145-153.

Tornatore, J. B., Hill, E., Laboff, J. A., & McGann, M. E. (2005). Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. *The Journal of neuropsychiatry and clinical neurosciences, 17*(1), 98-105.

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR mental health, 3*(2), e16.

Townhill, J., Hughes, A. C., Thomas, B., Busse, M. E., Price, K., Dunnett, S. B., . . . Rosser, A. E. (2016). Using Actiwatch to monitor circadian rhythm disturbance in Huntington' disease: A cautionary note. *Journal of Neuroscience Methods, 265*, 13-18. doi:https://doi.org/10.1016/j.jneumeth.2016.01.009

Trembath, M. K., Horton, Z. A., Tippett, L., Hogg, V., Collins, V. R., Churchyard, A., . . . Delatycki, M. B. (2010). A retrospective study of the impact of lifestyle on age at onset of huntington disease. *Movement Disorders, 25*(10), 1444-1450. doi:10.1002/mds.23108

10.1186/1471-2202-9-34

Tung, J. Y., Rose, R. V., Gammada, E., Lam, I., Roy, E. A., Black, S. E., & Poupart, P. (2014). Measuring life space in older adults with mild-to-moderate Alzheimer's disease using mobile phone GPS. *Gerontology, 60*(2), 154-162.

US Food and Drug Administration. (2017). *The voice of the patient: a series of reports from FDAs patient-focused drug development initiative*. Retrieved from

Vaccarino, A. L., Sills, T., Anderson, K. E., Endicott, J., Giuliano, J., Guttman, M., . . . Warner, J. H. (2011). Assessment of day-to-day functioning in prodromal and early Huntington disease. *PLoS currents, 3*.

Van Liew, C., Gluhm, S., Goldstein, J., Cronan, T. A., & Corey-Bloom, J. (2013). The functional implications of motor, cognitive, psychiatric, and social problem-solving states in Huntington's disease. *Psychiatry, 76*(4), 323-335. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84902301757&partnerID=40&md5=5e3f00f69e6c692491ee3f9521981ba9

Videnovic, A., Leurgans, S., Fan, W., Jaglin, J., & Shannon, K. M. (2009). Daytime somnolence and nocturnal sleep disturbances in Huntington disease. *Parkinsonism &*

*related disorders, 15*(6), 471-474. doi:http://dx.doi.org/10.1016/j.parkreldis.2008.10.002

Vidoni, E. D., & Burns, J. M. (2015). Exercise programmes for older people with dementia may have an effect on cognitive function and activities of daily living, but studies give inconsistent results. *Evidence-Based Nursing, 18*(1), 4. doi:10.1136/eb-2014-101737

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental Psychology*.

Walker, F. O. (2007). Huntington's disease. *The Lancet, 369*(9557), 218-228.

Walker, M. P., & Stickgold, R. (2014). Sleep, memory and plasticity. *Neuroscience and Psychoanalysis, 1*, 93.

Wallace, M., Downing, N., Lourens, S., Mills, J., Kim, J.-i., Long, J., & Paulsen, J. (2016). Is there an association of physical activity with brain volume, behavior, and day-to-day functioning? A cross sectional design in prodromal and early huntington disease.

Wesnes, K. A. (2014). Moving beyond the pros and cons of automating cognitive testing in pathological aging and dementia: The case for equal opportunity. *Alzheimer's Research and Therapy, 6*(8). doi:10.1186/s13195-014-0058-1

Wexler, N. S., Lorimer, J., Porter, J., Gomez, F., Moskowitz, C., Shackell, E., . . . Landwehrmeyeri, B. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proceedings of the National Academy of Sciences of the United States of America, 101*(10), 3498-3503. doi:10.1073/pnas.0308679101

Wiegand, M., Möller, A., Lauer, C., Stolz, S., Schreiber, W., Dose, M., & Krieg, J. (1991). Nocturnal sleep in Huntington's disease. *Journal of Neurology, 238*(4), 203-208.

Wiegand, M., Möller, A., Schreiber, W., Lauer, C., & Krieg, J.-C. (1991). Brain morphology and sleep EEG in patients with Huntington's disease. *European Archives of Psychiatry and Clinical Neuroscience, 240*(3), 148-152. doi:10.1007/BF02190755

Wild, K., Howieson, D., Webbe, F., Seelye, A., & Kaye, J. (2008). Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 4*(6), 428-437. doi:10.1016/j.jalz.2008.07.003

Williams, J., Downing, N., Vaccarino, A. L., Guttman, M., & Paulsen, J. S. (2011). Self reports of day-to-day function in a small cohort of people with prodromal and early HD. *PLoS currents, 3*.

Wood, N. I., Carta, V., Milde, S., Skillings, E. A., McAllister, C. J., Ang, Y. M., . . . Fernandes, J. X. (2010). Responses to environmental enrichment differ with sex and genotype in a transgenic mouse model of Huntington's disease. *PLoS ONE, 5*(2), e9077.

Worringham, C., Rojek, A., & Stewart, I. (2011). Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. *PLoS ONE, 6*(2). doi:10.1371/journal.pone.0014669

Yaffe, K., Falvey, C. M., & Hoang, T. (2014). Connections between sleep and cognition in older adults. *The Lancet Neurology, 13*(10), 1017-1028. doi:https://doi.org/10.1016/S1474-4422(14)70172-3

You, S. C., Geschwind, M. D., Sha, S. J., Apple, A., Satris, G., Wood, K. A., . . . Possin, K. L. (2014). Executive functions in premanifest Huntington's disease. *Movement Disorders, 29*(3), 405-409. doi:10.1002/mds.25762

Yuan, J., Li, M., Lv, G., & Lu, Z. K. (2020). Monitoring transmissibility and mortality of COVID-19 in Europe. *International Journal of Infectious Diseases, 95*, 311-315. doi:https://doi.org/10.1016/j.ijid.2020.03.050

Zhang, L., & Zhao, Z.-X. (2007). Objective and subjective measures for sleep disorders. *Neuroscience Bulletin, 23*(4), 236-240.

Zhang, Y., Ren, R., Yang, L., Zhou, J., Li, Y., Shi, J., . . . Tang, X. (2019). Sleep in Huntington's disease: a systematic review and meta-analysis of polysomongraphic findings. *Sleep, 42*(10), zsz154.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

Zorluoglu, G., Kamasak, M. E., Tavacioglu, L., & Ozanar, P. O. (2015). A mobile application for cognitive screening of dementia. *Computer Methods and Programs in Biomedicine, 118*(2), 252-262.

Zygouris, S., Ntovas, K., Giakoumis, D., Votis, K., Doumpoulakis, S., Segkouli, S., . . . Tsolaki, M. (2017). A preliminary study on the feasibility of using a virtual reality cognitive training application for remote detection of mild cognitive impairment. *Journal of Alzheimer's Disease, 56*(2), 619-627.

Zygouris, S., & Tsolaki, M. (2015). Computerized cognitive testing for older adults: a review. *American Journal of Alzheimer's Disease & Other Dementias®, 30*(1), 13-28.

# Appendix

## Appendix 1 Questionnaires included in HD-Mobile

Scale of Outcomes in Parkinson's Disease – Sleep (SCOPA-SLEEP) scale

*Presentation instructions*: Questions for this questionnaire will be answered on a discrete four-point scale, the response options are shown below. Participants will be presented with all options and will need to respond to all options before being able to continue.

***NS: Night time sleepiness***

Response options: never – sometimes – regularly – often

In the past *month,* …

1. … have you had trouble falling asleep when you went to bed at night?
2. … to what extent do you feel that you have woken *too often*?
3. … to what extent do you feel that you have been lying awake for *too long* at night?
4. … to what extent do you feel that you have woken up *too early* in the morning?
5. … to what extent do you feel you have had *too little* sleep at night?

Overall, how well have you slept at night during the past month?

response options: very well – well – rather well – not well but not badly - rather badly – badly – very badly

***DS: Daytime sleepiness***

Response options: never – sometimes – regularly – often

1. How often in the past month have you fallen asleep unexpectedly either during the day or in the evening?
2. How often in the past month have you fallen asleep while sitting peacefully?
3. How often in the past month have you fallen asleep while watching TV or reading?
4. How often in the past month have you fallen asleep while talking to someone?
5. In the past month, have you had trouble staying awake during the day or in the evening?
6. In the past month, have you experienced falling asleep during the day as a problem?

Hospital Anxiety and Depression Score

*Presentation instructions*: Questions for this questionnaire will be answered on a discrete four-point scale, the response options are shown below. Participants will be presented with all options and will need to respond to all options before being able to continue. **Note** If any participant endorses responses that put them in the "severe" category, the researchers will need to be notified immediately.

*Instructions for participants:* Please read each item below and select the reply which comes closest to how you have been feeling in the past week. Ignore the numbers printed at the edge of the questionnaire. Don't take too long over your replies, your immediate reaction to each item will probably be more accurate than a long, thought-out response

| | | |
|---|---|---|
| **A** | **I feel tense or 'wound up':** | |
| | Most of the time | 3 |
| | A lot of the time | 2 |
| | From time to time (occ.) | 1 |
| | Not at all | 0 |
| **D** | **I still enjoy the things I used to enjoy:** | |
| | Definitely as much | 0 |
| | Not quite as much | 1 |
| | Only a little | 2 |
| | Hardly at all | 3 |
| **A** | **I get a sort of frightened feeling as if something awful is about to happen:** | |
| | Very definitely and quite badly | 3 |
| | Yes, but not too badly | 2 |
| | A little, but it doesn't worry me | 1 |
| | Not at all | 0 |
| **D** | **I can laugh and see the funny side of things:** | |
| | As much as I always could | 0 |
| | Not quite so much now | 1 |
| | Definitely not so much now | 2 |
| | Not at all | 3 |
| **A** | **Worrying thoughts go through my mind:** | |
| | A great deal of the time | 3 |
| | A lot of the time | 2 |
| | From time to time, but not often | 1 |
| | Only occasionally | 0 |
| **D** | **I feel cheerful:** | |
| | Not at all | 3 |
| | Not often | 2 |
| | Sometimes | 1 |
| | Most of the time | 0 |
| **A** | **I can sit at ease and feel relaxed:** | |
| | Definitely | 0 |
| | Usually | 1 |
| | Not often | 2 |
| | Not at all | 3 |

| | | |
|---|---|---|
| **D** | **I feel as if I am slowed down:** | |
| | Nearly all the time | 3 |
| | Very often | 2 |
| | Sometimes | 1 |
| | Not at all | 0 |
| **A** | **I get a sort of frightened feeling like "butterflies" in the stomach:** | |
| | Not at all | 0 |
| | Occasionally | 1 |
| | Quite often | 2 |
| | Very often | 3 |
| **D** | **I have lost interest in my appearance:** | |
| | Definitely | 3 |
| | I don't take as much care as I should | 2 |
| | I may not take quite as much care | 1 |
| | I take just as much care | 0 |
| **A** | **I feel restless as I have to be on the move:** | |
| | Very much indeed | 3 |
| | Quite a lot | 2 |
| | Not very much | 1 |
| | Not at all | 0 |
| **D** | **I look forward with enjoyment to things:** | |
| | As much as I ever did | 0 |
| | Rather less than I used to | 1 |
| | Definitely less than I used to | 2 |
| | Hardly at all | 3 |
| **A** | **I get sudden feelings of panic:** | |
| | Very often indeed | 3 |
| | Quite often | 2 |
| | Not very often | 1 |
| | Not at all | 0 |
| **D** | **I can enjoy a good book or radio/TV program:** | |
| | Often | 0 |
| | Sometimes | 1 |
| | Not often | 2 |
| | Very seldom | 3 |

| Score | Interpretation |
|---|---|
| 0 – 7 | Normal |
| 8 – 10 | Mild |
| 11 – 14 | Moderate |
| 15 – 21 | Severe |

## Daily Mood, Sleep and Activity Questions

Each day participants will be asked to answer a few brief questions regarding their mood, sleep and activity information from the previous day. The questions are:

1. For how long (if at all) did you nap yesterday?
2. At what time last night did you close your eyes with the intention of falling asleep?
3. After deciding to go to sleep, how many minutes did you take to fall to sleep?
4. How many times did you wake up during the night?
5. Adding all the time you woke, how many minutes did you spend awake during the night?
6. What time did you wake up this morning (for the final time)?
7. What time did you get out of bed this morning (for the final time)?
8. Overall, how well did you sleep last night (7-point likert scale, very well to very badly)
9. In the previous day, how often did you feel tense or wound up? (four-point likert scale, not at all to most of the time).
10. In the previous day, how often did you feel cheerful? (four-point likert scale, not at all to most of the time)
11. Thinking back to yesterday, in the course of your day, approximately how long did you spend doing moderate or vigorous physical activity? This can include work, leisure, household and garden activities. Vigorous physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. Moderate activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal).

12. Thinking back to yesterday, in the course of your day, approximately how long did you spend sitting? This may include time spent sitting at a desk, visiting friends, reading or sitting or lying down to watch television, or time spent sitting in a motor vehicle.

## Karolinska Sleepiness Scale

Please mark the number that best corresponds to how sleepy you feel right now. You may mark any number, but mark only one number.

_____ 1. Extremely alert

_____ 2. Very alert

_____ 3. Alert

_____ 4. Rather alert

_____ 5. Neither alert nor sleepy

_____ 6. Some signs of sleepiness

_____ 7. Sleepy – but no difficulty remaining awake

_____ 8. Sleepy, some effort to keep awake

_____ 9. Extremely sleepy – fighting sleep

**Appendix 2 Semi-structured Interview Script**

How do you rate the overall useability of the CogLab Mobile App (1 – 10)

How confident did you feel in being able to understand how to complete the questionnaires? (1 -10)

How confident did you feel in being able to understand how to complete the Tapping Task (Brain Game 1)? (1-10)

How confident did you feel in being able to understand how to complete the SnowflakesTask (Brain Game 2) (1-10)

How confident did you feel in being able to understand how to complete the Memory Task (Brain Game 3) (1-10)

Overall, how did you feel about completing tasks and questionnaires on your smartphone?

Do you have any suggestions for improving CogLab mobile

How useful/informative did you find the video instructions? (1 – 10, N/A did not watch)

How Useful/informative did you find the written instructions (1 – 10, N/A did not read)

Following reading/viewing the instructions how confident did you feel that you could complete the study independently? (1-10)

Do you have any suggestions for improving the participant instructions?

How do you rate the overall usability of Fitbit One? (1 – 10)

How confident did you feel in setting-up and using the Fitbit (1-10)

Did you remember to wear the Fitbit at all times?

Did you remember to set sleep mode every night?

Did you encounter any problems/difficulties during your participation?

# Appendix 3 Supplementary Material for Study Three

**Brendan McLaren** **Ricardo J. Romeu** **Sophie Andrews** **Mark Bellgrove**

**Julie Stout**

## Introduction

Here we present the details of the two main classes of models that we tested against the data reported in the main text. Both are hierarchical diffusion models that use the 'dwiener' package in JAGS (Vandekerckhove, Tuerlinckx, & Lee, 2010), as discussed in the main text; however, they differ in the assumptions that are made. We will also highlight, when appropriate, minor variants that were also tested.

The first model, which we call the *variability* model, is a basic diffusion model, largely following the example given in (Vandekerckhove et al., 2010); however, the variability model will also treat variance in mean drift rate between the three groups as a parameter to be estimated. We expand upon this more below.

The second major model is a *regression* model, wherein we suppose that we can regress the mean drift rate on the age of the subject, and the intercept, slope, and variance of this regression equation will depend on group (for one variant) or on subject, which comes from a group (for the second variant; see below). These models were tested because age is a known factor that can affect response times (Ratcliff & McKoon, 2008), and since our main focus was on the drift rates, we naturally thought to regress on age of the participants to see if there was an effect above and beyond that of merely aging. While this may be a helpful exercise from a diffusion modeling perspective, we should note there may be an objection to regressing on age from a Huntington's researcher's perspective. The main objection is that the disease is inexorably linked to aging, as it is a progressive disease that naturally becomes worse with age (Walker, 2007). Therefore, it could be that regressing on age may actually remove the very object of study for a Huntington's researcher. Nonetheless, as

we will see, the regression models largely agree with the results presented in the main text, and in many instances even give us the same estimated values for the parameters across groups.

All code used to fit the models, including the text files with the JAGS models we will detail in a moment, can be found on the Open Science Framework page corresponding to this article: https://osf.io/k2urm/.

Before we begin with the model details, however, we will briefly review the basics of Bayesian modeling using a simple coin flipping example. While this will certainly not suffice for the complete novice to Bayesian analysis, we hope the simplicity of the example will help illustrate the main concepts that are important for interpreting and understanding the main modeling results. More information on Bayesian modeling can be found in (Kruschke, 2014; Wagenmakers et al., 2018; Vehtari, Gelman, & Gabry, 2017).

## Bayesian Modeling

The most fundamental aspect of Bayesian modeling is to explicitly state and incorporate our prior assumptions about a set of data we wish to model, in such a way that our observed data may change and update our prior beliefs in a reasonable way. The end result - that of transforming our prior beliefs by observing data - is known as a *posterior* distribution, and these distributions will be of great importance when analyzing our model parameters.

To illustrate, we take a modified example from Kruschke (2014). Suppose we are flipping a coin and we wish to determine its probability of it coming up "heads." This probability is a latent variable that must be inferred from observing the coin being flipped, and so we perform a simple experiment wherein we flip the coin, say, $N = 30$ times and record $z$, the number of heads observed. If we have no information on the coin, we might suppose that any probability of coming up heads (call it $\theta$) is equally likely. We can represent this *prior belief* by placing a $Beta(1, 1)$ distribution over the possible values of $\theta$ – this can be seen as the blue curve in the Fig. 1. Now, suppose we have flipped our coin 30 times and happened to have observed $z = 23$ heads. What should our posterior probability over $\theta$ be? What makes this example so simple is that we can arrive at a simple formula for calculating the posterior distribution over $\theta$: our posterior will be $Beta(a_0 + z, b_0 + (N - z))$. That is, we add the number of heads to the first parameter $a_0$ and add the number of tails to the second parameter $b_0$. In our case, $a_0 = b_0 = 1$. Thus, our posterior is $Beta(1 + 23, 1 + 7)$, and this curve is shown in black on Fig. 1.

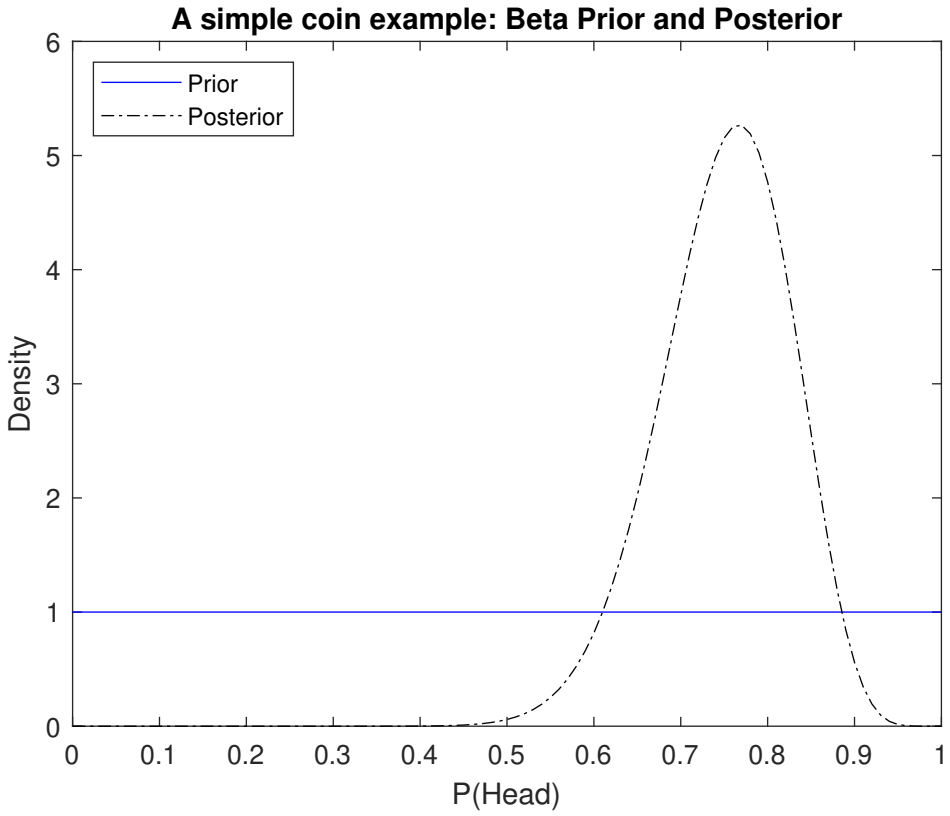



Figure 1: An illustration on how our prior beliefs should change in accordance with observed data, using Bayes' Rule.

The reader should immediately take note that the mean of the posterior is around .75, indicating that the data lead us to believe the coin to be biased. This should make perfect sense, given the particular data sequence we have observed. While this example is extremely simple, the basic ideas still pertain to the more complicated diffusion model we will discuss in due time: (1) we need to explicitly establish our prior beliefs in the form of a distribution over the admissible values of our parameter(s) of interest; (2) we need to input the observed data into the model, yielding us a posterior estimate of each parameter. Note that the estimates we obtain are *not* point estimates, but rather an entire distribution of possible scores. Indeed, even in our simple coin example, our model recognizes that $\theta$ values other than .75 are possible, just not as likely. That is, the posterior distribution captures not only what values are most likely, but also details the amount of uncertainty we have in that estimate after having observed our data.

Of course, in more general settings, the models become more complicated, and as a result, the Bayesian analyses also become more complicated. In general cases,we do not have an explicit and simple formula for arriving at our posterior distributions, and we must therefore estimate them

through simulation. Programs like JAGS make it easy for the general researcher to simulate from the posterior distribution - which gives one a set of samples that can be used to estimate most properties of the posterior with ease - without having to hand-code these simulation techniques (which are known as *Markov Chain Monte Carlo* [MCMC] simulations). While the actual implementation becomes a bit more complicated, the same basic principle of updating priors in light of new data is the same.

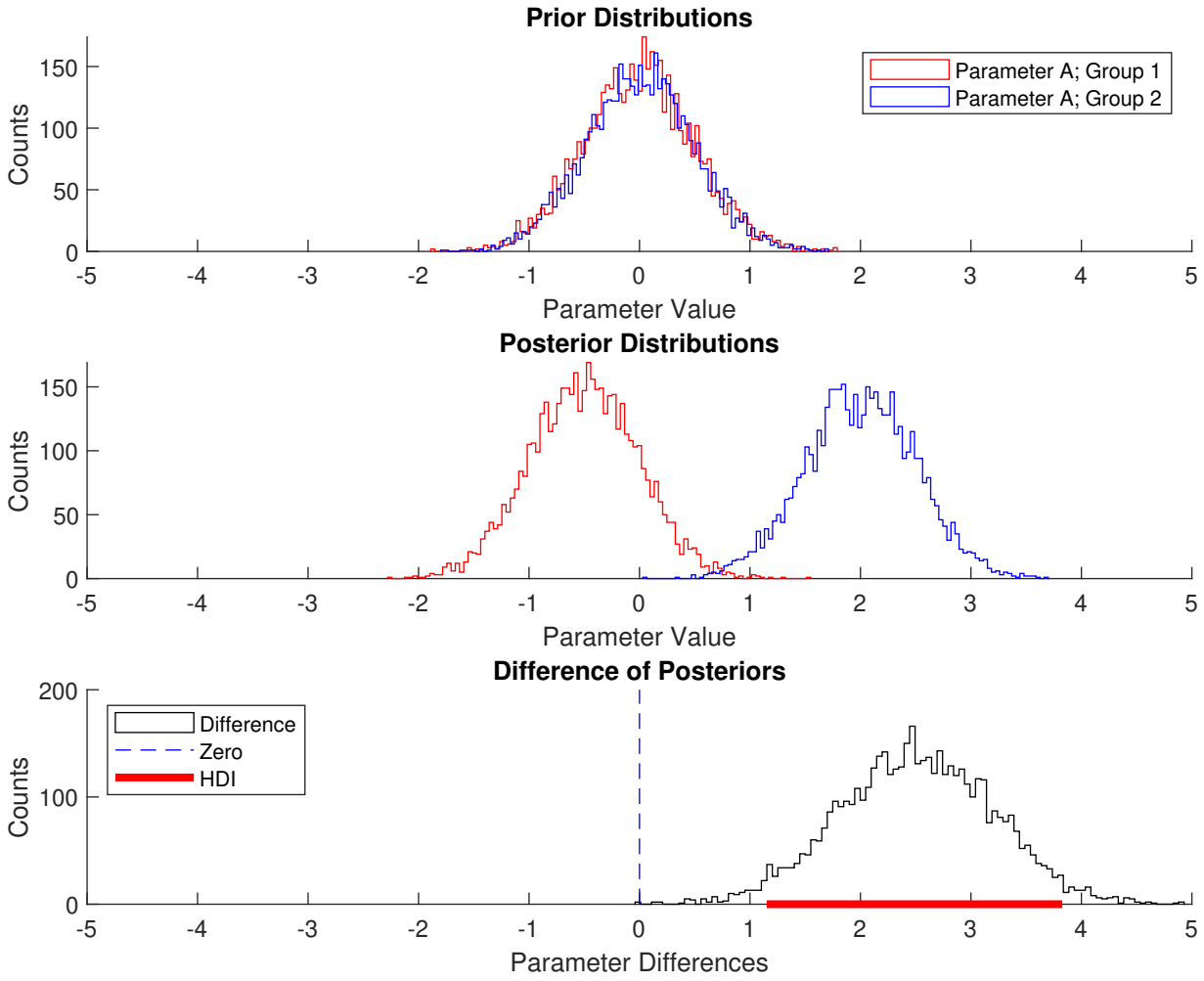



Figure 2: An example of updating beliefs about a parameter difference between two hypothetical groups. The top panel shows the priors, the middle shows the posteriors after having observed data from both groups, and the final panel shows the difference between the two posteriors (Group 2 - Group 1). See the text for more details.

In particular, MCMC methods will give us sampled points from the posterior distribution, and these samples will accurately reflect the true posterior distribution if the MCMC chain is allowed

to sample for a long enough time (Kruschke, 2014; Wagenmakers et al., 2018). Indeed, plotting a histogram of the MCMC samples gives a rough estimate of the true posterior distribution. We can then ask essentially any question of interest by probing these posterior samples. For example, in relation to our coin example, we might ask: (1) What is the probability of the bias in the coin being greater than or equal to .75? This would be estimated by noting the proportion of posterior samples that are at or above .75. (2) Is the coin a fair one? We can give an estimated answer to this by observing the *highest density interval* of the posterior samples, which is similar in concept to a confidence interval for the posterior distribution. However, the highest density interval actually gives us the range of values that captures 95% of the sampled values, so that the true value has a 95% chance of being within the highest density interval (once we accept the observed data, that is). If the coin is biased, then the highest density interval should exclude .5 as a credible (i.e., likely given the data) value for the coin. Applying both of these questions to our posterior samples will generally answer most of the questions we might have about a model, and are in fact the main questions we asked about the best fitting diffusion model in the results section of the main paper.

Perhaps a concrete will be helpful for the naive reader. Figure 2 presents an example of Bayesian updating in action; the jagged histograms are there to remind the reader that these distributions are not exact, but rather a plotting of sampled points that are meant to *approximate* the true, underlying posterior. The figure overall represents a typical Bayesian hypothesis test using HDIs. This example, e.g., represents the prototypical scenario involved in testing if two groups have a different value for some parameter in a model. At the top of the figure, we see the prior distributions of the parameter for both groups; they overlap because this example assumes the same prior distribution for each group. The middle panel depicts the *estimated posterior distributions* of the parameter for each group; the reader can see that the distributions have now separated, indicating that, with new evidence, we no longer believe that both groups have a similar distribution for this parameter. The bottom panel shows the *posterior difference distribution* between the two tested groups for this particular parameter; this distribution is obtained by "subtracting" one distribution from the other. To better understand this, remember that each depicted histogram is represented by a vector of sampled points from the posterior; to obtain the difference distribution, we simply subtract one vector from the other and plot the resulting histogram. Using the difference distribution, we can see what the posterior difference value is relative to a difference of zero between the two groups; zero is represented as the vertical line in the bottom panel. The HDI is represented as the horizontal, red line on the x-axis that lies below the bulk of the posterior distribution. This red line shows

the interval within which there is a 95% chance that the true value of the difference is within that interval, given the data and the model used. The expected difference, given the data and model, can be obtained by taking the mean of this histogram. The probability of a difference, say Group 2 > Group 1, can be found by finding the proportion of this difference histogram that is above zero. (In this case, we look at *above zero* because this difference happens to be Group 2 - Group 1, so that $G_2 - G_1 > 0$ is the same as $G_2 > G_1$.)

From here, the Bayesian hypothesis test would check if the hypothesis of a zero difference is a credible one, given the model and the data. This is checked by observing if zero is within the HDI of the difference distribution; this is because the HDI indicates the most likely parameter values, so being outside the HDI is equivalent to being a highly unlikely value, given the data[1]. In this example, zero is well outside the HDI, so that we would conclude that we have credible evidence that groups 2 and 1 differ on this particular parameter; looking at the proportion of the histogram above zero will give you the probability that the difference exists, while the mean and standard deviation of the differences will give an estimate of the expected difference, plus the range of expected noise.

## The Variability Model

In this section, we will detail the modeling assumptions and implementations that were used for the main model presented in the text, as well as some variants that we tested alongside the main-text model (which in this supplement we will call the *variability* or *variable* model for reasons that will soon become apparent). The reader is encouraged to compare the following exposition with the JAGS model in the "VariabilityModel.txt" file.

First, we input the data into JAGS using the "y" vector, which is a column vector that stores all the response times across all subjects and conditions (with the total number of data points given as Ntotal). Each element of $y$ is a response time in seconds, and is positive for correct trials and negative for incorrect trials. All response times less than 100 ms were discarded from analysis, as these response times are generally considered too fast to be "real" responses, and can severely impact the estimation of $\tau$, the non-decision time, among other parameters. By "discarded," we mean that these data points were recoded as missing data, and MATLAB is told this by making the corresponding element in $y$ a *NaN* value (for Not A Number). The response times in $y$ are distributed

[1]It is important to emphasize that the results are dependent on the particular model that was chosen for analysis *and* the observed data. So as to prevent a repetitive "given the data and model" throughout, the reader should keep in mind that all statements related to a Bayesian hypothesis test, including point estimates from the posteriors, are assuming a given model and data set.

according to the dwiener distribution, which is dependent on boundary ("alpha"), non-decision time ("tau"), bias ("beta"), and the drift rate ("delta"). The reader should note that alpha is dependent on the condition ("cond") and the subject ID ("subj"), while the non-decision times (NDT), the biases, and the drift rates are changing from trial-to-trial (i.e., only indexed by $i$, which runs through all data points). This is implementing the trial-to-trial variability of these parameters, an assumption that has been adopted by the latest developments of the diffusion model (Ratcliff & McKoon, 2008; Ratcliff et al., 2016). Following the data vector $y$, we see that we make the assumption that: (1) $\tau$ (NDT) comes from a uniform distribution with some upperbound $U_\tau$, which is itself assumed to be a subject-level parameter; (2) $\beta$ (bias) is assumed to come from the Beta distribution, with subject/condition level parameters of "Abeta" and "Bbeta", meaning that we assume the bias can change from person to person, and also from condition to condition within a single person; and (3) $\delta$ (the drift rate on the current trial) is assumed to come from a normal distribution, with a mean $v[c, s]$ dependent on the condition (Few vs Many) and the subject ID, and a precision[2] of 1. We emphasize that these last 3 parameters are indexed by $i$, meaning that they change from trial to trial.

Next come the specifications of the subject-level parameters. These are the "individual differences" parameters whose mean estimates were used in the correlational analyses in the main text. Remember that the Bayesian estimation procedure of MCMC will output a distribution of possible scores; we can extract a point estimate by merely taking the mean of this distribution. First is the upperbound $U_\tau$, which is assumed to be distributed as an exponential with some rate $\eta_g$ that is dependent on the group of the subject. This seemed a reasonable guess at a prior for $U_\tau$, since NDTs tend to be in the range of 100 - 300 ms or so, and since our data are coded in seconds, this translates to small values for the NDT. Note that the NDT is not assumed to change across conditions. The subsequent "for-loop" traverses the 2 conditions in the experiment, and thus all parameters within this condition loop are dependent on the subject and the condition. The next line says that the subject-level mean drift rates (denoted as $v$) are distributed as normal with a mean $\mu_{c,g}$ dependent on the condition and the group from which subject $s$ originates, and with a precision estimate "prec" that is also dependent on the condition and group. *Indeed, this "prec" random variable is what gives the variable model its name,* as the typical assumption is to set this standard deviation variable to a value like .1 or 1 (Ratcliff  McKoon, 2008). While setting the variability is common, this makes the assumption that the variances are the same across all groups, and we wanted to test that assumption by estimating the variance in each group. Next comes the boundary parameter $\alpha$, which is assumed

[2]JAGS parameterizes the normal distribution as having a mean and a *precision*, which is given by $1/(\sigma^2)$, where $\sigma$ is the standard deviation.

to come from a Gamma distribution (it can only take positive values here), where the rate and shape parameters "Aalpha" and "Balpha" are assumed to change between conditions and groups. It is from these rate and shape group-level estimates that we can compare the boundaries across groups. A Gamma distribution was chosen for its support being non-negative real ($\geq 0$) and being a fairly flexible distribution. We also assume that the shape parameters from the Beta distribution for the bias (Abeta and Bbeta) each came from a Gamma distribution, and each of these distributions has a condition/group level rate and shape parameter. The Gamma distribution for Abeta and Bbeta were chosen for similar reasons to the boundary choice.

Finally come the group-level parameters, where the hierarchy stops. Most of these parameters are assumed to change between conditions and groups, though the rate of the upperbound for the NDTs was assumed to only change across groups. From the first line under the "# Group level priors:" section, we assume that $\eta[g] \sim$ Gamma$(1, 1)$, a somewhat uninformed prior[3]. Under the subsequent condition loop within the group-level parameters, we see that the mean of the mean drift rates $\mu[c, g]$ was assumed to come from a standard normal distribution. The next line converts the standard deviation (the variability in the model) to a precision, and the standard deviation is assumed to come from a Gamma(1,1) distribution (must be positive). The "Aalpha" and "Balpha" were assumed to come from an exponential distribution, as it was not expected that the boundaries would be very large. Finally, the rates and shapes of the Gamma distributions for the subject-level bias parameters Abeta and Bbeta were all assumed to be themselves drawn from a Gamma(1,1) distribution. We observed group differences by examining the posterior distributions of all of these parameters (or some combination to look at means, medians, etc.).

There were two major variations on this main model, the first we arbitrarily called the "Standard model" (because it was the first that was tested) and the other we called the "Alternative model," because some of the priors were an alternative to the ones chosen initially.

In particular, in the Standard model ("HierHDmodel.txt"), both the non-decision time and the bias were assumed to come from a uniform distribution, each with its own upperbound ("Ut" and "Ub", for the NDT and bias, resp.). They were both assumed to be only dependent on subject and not on condition. "Ut" was assumed, for each person, to come from an exponential distribution with some rate $\eta_g$ dependent on group. "Ub" was assumed to come from a Beta distribution, as it must be within the range of [0, 1] (recall that the bias was parameterized as a percentage of the

[3]In fact, all the priors for the group-level distributions were chosen to be relatively uninformed, as it was not clear exactly what these parameters should look like.

boundary). This subject-level Beta distribution had its own parameters "Au" (A for upper) and "Bu" (B for upper). Each of these were only dependent on group, and, for each group, they were assumed to come from a Gamma(1,1) distribution. The other notable change here is the prior distribution for $v[c, s]$, the mean drift rate for each individual. It still comes from a normal distribution, with an overall mean dependent on group and condition, but the precision here is now a constant of 1, not estimated. Indeed, it is a fixed constant of 1 *for all groups*.

In the Alternative model ("AltHierHD.txt"), we changed largely the prior distributions of some variables and their dependency on condition. We assumed that the trial-to-trial bias "beta[i]" comes from a Beta distribution, in exactly the same way as in the Variable model, with Abeta and Bbeta as before. Unlike in the standard model, however, the Alternative model allowed the bias to change between condition and subject, which makes the Alternative model have more parameters than the Standard model.

It should also be noted that the prior for the NDT was also changed in some implementations of these models, in such a way as to bound the NDT away from zero. In fact, we had an implementation that was similar to that in Ratcliff et al. (2016) and in Ratcliff & McKoon (2008), where the NDT comes from a uniform distribution centered around some value $T_z$ and with some spread $s_z$. However, our programs had a great deal of trouble implementing this model, because it was always possible to sample values for the NDT that were larger than most of the available response times. (In fact, it was sometimes very likely, depending on the starting values of the Markov chains.) Indeed, bounding the NDT away from zero always had the possibility of sampling a value that was greater than some of the response times in the data, and sampling such a value made the model invalid, and would thus make the program crash. Even having a more liberal criterion for removing response times did not seem to alleviate this problem. The estimation of the model seemed to always be subject to randomly crashing, making the implementation of the models unreliable in practice. We implemented the uniform distribution because it had been used in previously successful work with the dwiener module in JAGS (Vandekerckhove et al., 2010); however, as noted in the main text, this led to extremely unreliable estimates at the subject-level. Future work should experiment more with this issue, since at this point, it is unclear if the major problem is the ambulatory nature of the data (few data points, many missing data points sometimes) or if the problem is the assumption of a uniform distribution for the NDTs.

Table 1 demonstrates the DICs (Deviance Information Criterion - a measure of model fit to the data) for each of the models tested.

| *Model* | *DIC(Var) - DIC(Model)* |
|---|---|
| Standard | -293.7 |
| Alternative | -183.7 |
| **Variability** | 0 |
| **Regression**-Subject | 306.3 |
| **Regression**-Group/Condition | 89.7 |
| **Regression**-Group Only | 3.2 |

Table 1: The (relative) Deviance Information Criterion for each of the models tested on the data, relative to the *variability model* for ease of interpretation. Negative values indicate a worse fit than the variability model, while positive values indicate a better fit. While the regression models generally fit better than the variability model, they produce the same parameter estimates, albeit through "noisier" posterior distributions. The regression models treated the group-level mean drift rates as affine functions of the age of the participant. We focus on the variability model for ease of presentation and its easier interpretation, but highlight some caveats in the section of this supplement related to the regression models.

## The Regression Model

The regression model is, in many ways, identical to the variability model. That is, it has many of the same prior distributions, using the same justifications of a lack of prior knowledge, and similarly is hierarchical, allowing for an assessment of group differences as well as individual differences. The main difference between the regression model and the variability model is the allowance of each individual mean drift rate $v[c, s]$ to be described thus:

$$\begin{aligned} v[c, s] &\sim \text{Normal}(\mu[\text{group}[s]], \epsilon[c, \text{group}[s]]) \\ \mu[\text{group}[s]] &= \beta_0[\text{group}[s]] + \beta_1[\text{group}[s]] * \text{Age}[s]. \end{aligned} \tag{1}$$

That is, each individual mean drift rate $v$ comes from a hierarchical group-level normal distribution, with this group distribution having a mean given by the simple linear regression equation in (1). This regression equation states that the intercept and slope are group-dependent only, and the mean of the mean drift rate $v$ is determined by these parameters applied to the added data of *age* for each person. Note that in this particular model, the regression on the mean of means is not dependent on the condition $c$ of the experiment; only the variance is assumed to change across conditions.

Another instance of the regression model that we implemented allows the regression coefficients to change across each subject, while also allowing the variance of $v[c, s]$ to change across groups and conditions:

$$\begin{aligned} v[c, s] &\sim \text{Normal}(\mu[s], \epsilon[c, \text{group}[s]]) \\ \mu[s] &= \beta_0[s] + \beta_1[s] * \text{Age}[s]. \end{aligned} \tag{2}$$

Thus, the regression equation is similar to before, but now allows the coefficients to change according to each subject. It is additionally assumed that each intercept parameter $\beta_0[s]$ and each slope parameter $\beta_1[s]$ for each subject $s$ is drawn from a group-level distribution as follows:

$$\begin{aligned} \beta_0[s] &\sim \text{Normal}(\mu_{Int}[group[s]], 1) \\ \beta_1[s] &\sim \text{Normal}(\mu_{Slope}[group[s]], 1). \end{aligned} \tag{3}$$

Both of $\mu_{Int}[g]$ and $\mu_{Slope}[g]$, for each group $g$, was given a hyperprior of coming from a standard normal distribution.

As the DIC's demonstrate, the regression models, especially the subject-level regression model, performed better than the non-regression models. However, their interpretation is a bit cumbersome, and in fact yields largely the same estimates as the variability model presented in the main text, which is why we decided to focus on the variability model.

Figure 3 presents the regression lines for each group, as estimated from the mean of the posterior distributions for each regression coefficient using the "Group Only" model for simplicity. While these regression lines increase with age (which might not be expected), this may be somewhat misleading, as the posterior distributions had a large amount of variability relative to the estimated *a posteriori* means. More data are needed in order to confirm the increasing slope of these regression lines (an example of the noisy posterior distributions for the intercepts is given in Figure 4). Nevertheless, taking the coefficients as valid, we see that the three groups become more and more disparate in their mean drift rates as age increases. This suggests that a Huntington's diagnosis has an effect on encoding of stimuli above and beyond that of merely aging.

Figure 5 presents estimated distributions of drift rates for all three groups. These distributions represent the distributions of individual drift rate estimates for each population; they are estimates of where the $v[c, s]$ come from for each subject, given their group membership. These estimates were calculated by using the group-regression model estimates for the regression coefficients, and using the average age of each group in the regression equation. The standard deviation was estimated from the group level variability in drift ($\epsilon$) for each group. The distributions, especially those for the Huntington group and the healthy controls, are fairly separable. Indeed, these separations are replicated in the analyses for the variability model. However, the regression model found no appreciable difference in drift rates across conditions for any of the groups, which suggests that the within-group difference in mean drift rate for healthy controls in the main text is probably

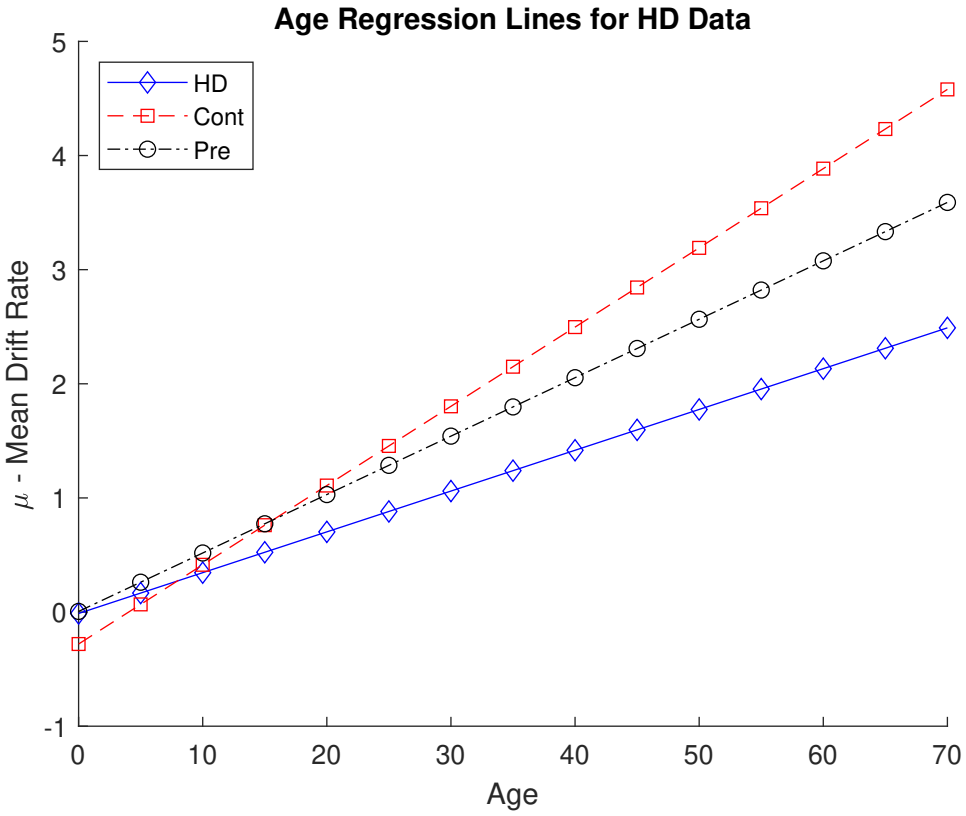



Figure 3: Regression lines for each group of mean drift rate against age. There appear to be clear separations between the groups as age increases. The mean age for the premanifest and healthy control groups were about 44, and the mean age for the Huntington's group was about 50.

attributable to the spread in ages in the group. This is to be expected, as this is how the model is coded, but additionally we did not find any appreciable difference in variance across the two conditions, and variance was allowed to vary freely across groups and conditions. It should also be emphasized that the means of these distributions (i.e., the means estimated by the regression lines) are very similar to those estimated using the variability model (i.e., with no regressing on age).

## Limitations of the Model

While the model was able to demonstrate expected (and hoped for) group differences, there are some limitations to the model that should be kept in mind. First and foremost, as these analyses were done on ambulatory data, there are a large number of missing data points involved in the analyses, and the number of known data points per person were much less than is typically collected for a diffusion model analysis (Ratcliff & McKoon, 2008). While the Bayesian method was implemented to counteract these difficulties, they should nonetheless be kept in mind when evaluating the evidence.

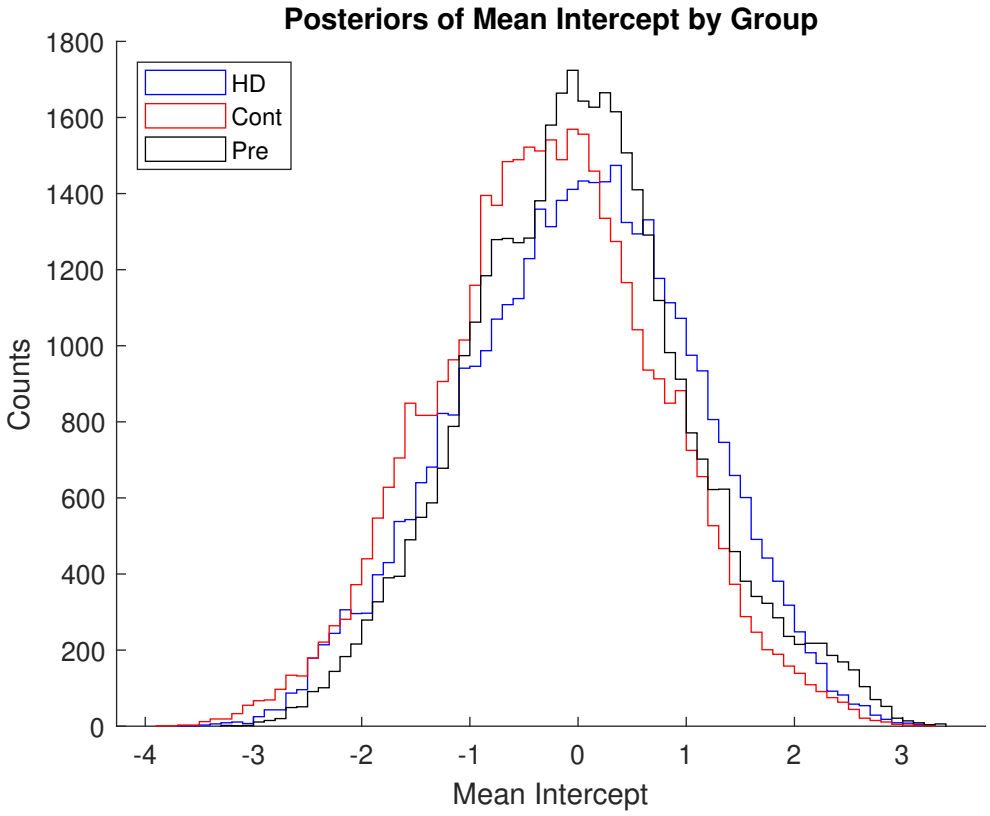



Figure 4: An example of comparing posterior distributions of regression coefficients from the group-regression model. As can be readily seen, the variance in these estimates is quite large, and the means are all close to zero, suggesting that more data are needed in order to have more confidence in the estimates. A similar case occurs for the estimation of the slopes.

Certainly, more data is better, especially in the Bayesian framework, where our posteriors could serve as the interested readers' priors (Kruschke, 2014; Wagenmakers et al., 2018).

A more pressing limitation is the lack of results for the non-decision time in the main models we tested. The major issue was that convergence of the MCMC chains for the individual-level non-decision times ($\tau$) was very often absolutely horrendous. As mentioned in the main text, these convergence issues mean we cannot place a great deal of trust in the estimates, as the posteriors estimated in this fashion are not accurate reflections of the underlying true posteriors. An attempt to mitigate this problem was the use of the *Alternative* model, which changed the prior distribution on the non-decision times. However, this model also suffered with its $\tau$ estimates. Indeed, all the models tested suffered this issue. This is a shame, because there are known motor issues with Huntington's, and it would be a great validation of the model if it were able to capture that expected difference. While the group-level estimates of the non-decision times had good convergence statistics, these estimates were undoubtedly affected by the poor estimation of the individual level

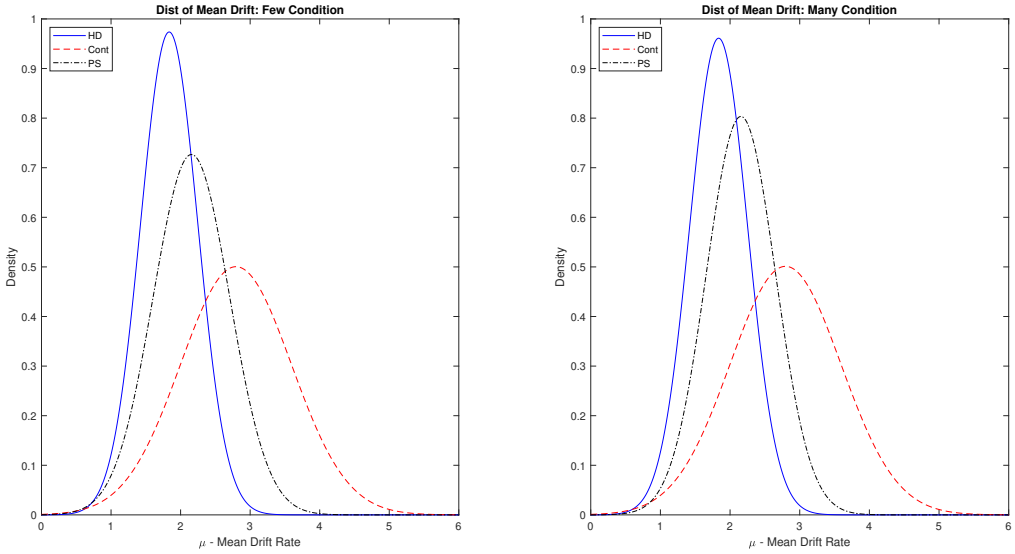



Figure 5: Posterior mean estimates of the group level distributions of drift rate. That is, these are the distributions: $v[c, s] \sim \text{Normal}(\bar{\beta}_0 + \bar{\beta}_1 \bar{A}, \epsilon[c, \bar{\text{group}}[s]])$, where the bars indicate a sample average. There are readily apparent separations that are more easily tested using the variability model from the main text. The means of these estimated posteriors are surprisingly similar to the means obtained without regressing in the variability model. The left panel has the posteriors for the Few condition, the right for the Many. The means are, by assumption, not different, but the variance is very slightly different between the two.

estimates; thus, we should treat the observed null group level results of the non-decision times with caution. We speculate that it is the ambulatory setting of the data, including both the limited number of data points and the (sometimes large) number of missing data points per person (more on this below), that is the major culprit of this estimation problem. It may be that more data are needed to accurately estimate this parameter at the individual level, even in a hierarchical Bayesian framework. Future work could find better implementations of the non-decision time in the hierarchical model so that it is estimated accurately and robustly in such an applied setting.

## Extra Details

In this section, we detail some extra insights that could not fit into the main text. We go over the details of the missing data, with particular focus on the amount for each group and a discussion of the consequences of this for the model; we also highlight the reliability issues we ran into - for the non-decision time in particular - for the variability model; and finally, we touch on an interesting consequence of the variability estimates in our model when wanting to practically implement the model in clinical work.

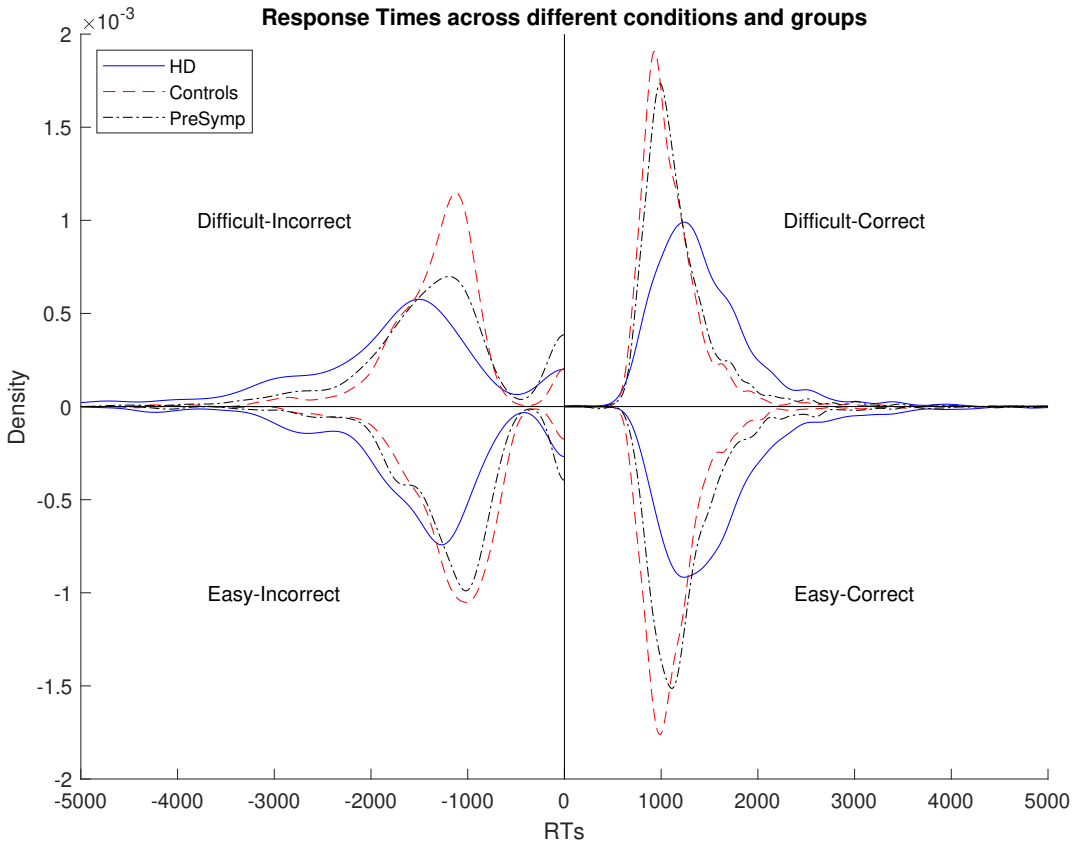



Figure 6: Group-level estimated density kernels for the response times of each group, partitioned by condition ("few"/"easy" vs. "many"/"difficult") and correctness (correct/incorrect). Response time is given in milliseconds. Overall, we can see a clear rightward shift (higher response times in the median/average) in the HD group compared to the controls in particular, and possibly in the pre-manifest group as well, along with fatter tails in the HD group. This latter result indicates that the HD group is much more likely than the other groups to have longer response times.

Before we do so, however, we present Figure 6 that depicts the response time distributions for each group, partitioned by both condition ("few" vs. "many"; "easy" vs. "difficult") and correctness. It seems clear from the figure that we should expect a significant different between the HD group and the controls; the results involving the pre-manifest patients is a bit more subtle, but nonetheless captured by the models, as described in the main text. The reader should also note the fast errors, i.e., the "bumps" that occur to the left of zero in all groups and in both conditions. Overall, at least with respect to the clear difference between HD and controls, it appears that the model needed to explain both a *rightward shift* (i.e., longer response times) in addition to a *fattening of the tails* of these distributions, in both conditions. (The fatter tails can be seen by noting that the estimated density for HD dominates the density for the other two groups after approximately 1200

ms in all correctness and condition combinations.) From the results of the main text, the model suggests that a lower mean drift rate in the HD group, combined with a somewhat smaller spread of individual differences of drift rate within the HD group, was sufficient to explain both of these apparent phenomena in these response time distributions. Indeed, a smaller drift rate would push the distribution "rightward" (i.e., increase response times, on average/median), and with the group as a whole a bit more homogeneous in their lower drift rates than controls, we might expect a higher chance in the HD group of having much longer response times, leading to the fattened tails for the HD group.

In addition, the reader should note that the response time distributions for the pre-manifest and control groups are nearly identical for correct response times, but there is a slight rightward shift and fattening of tails for *incorrect* trials; this is especially pronounced in the incorrect response times in the "many"/"difficult" condition.

**Missing Data**

| Group | Total Data | Missing Data | Max Proportion |
|---|---|---|---|
| HD | 240 (26) | 5.8 (12) | 15% |
| Premanifest | 258 (60) | 6.9 (14) | 13% |
| Controls | 237 (15) | 1.8 (5) | 10% |

Table 2: **Analysis of Missing Data.** Number of total trials, with average (standard deviation), for each group, along with the average number of missing trials (standard deviation) for each group. The "Max Proportion" statistic finds the maximum proportion of missing/total trials to give a more relative measure. Overall, we found that, while the clinical groups tended to have more total trials, they also tended to have more missing data on average, and greater variability in the amount of missing data per person. Indeed, this is corroborated by the higher max proportions in the clinical groups (higher missing:total ratios in the clinical groups).

Table 2 contains the basic statistics related to missing data from each group. Each person could vary in the number of total trials, because they may differ in the number of days in the assessment that chose to participate in, as well as in the number of missing data points recorded for each person. The most common reason for missing data was a response that was too quick. There were also a couple participants that appeared to start the task normally, but then rush through when they got bored, as indicated by having starting data similar to the previous session, followed by a long string of missing data until the session was over. As the Bayesian approach allows for the imputation of these values from past data, we left these missing values in our model fits, especially given that, even with a complete set of data, the number of trials is typically quite small for a diffusion analysis, as mentioned in the main text.

While a natural imputation method, which would not rely on JAGS, would be to replace the missing data with the mean of the complete data, the worry was that this approach would have a strong influence on the estimation of all parameters. Indeed, the relatively high amount of missing data in the clinical groups, combined with an already low maximum amount of data, would have almost certainly led to this mean-imputation method giving us tainted results in all parameters from the model, as having the exact same repeated value in response times would make the distribution the model is trying to fit a degenerate one. This would lead either to a complete breakdown in the estimation process, or in estimates that are completely nonsensical. Letting JAGS handle the imputation, then, seemed like a reasonable compromise.

**Reliability and Stability of the Variability Model Parameter Estimates**

As mentioned in the main text, we ran into some issues of stability of our estimates. The exact details can be found in the relevant text files posted to the article's OSF page ("RhatSummary2.txt"). While the group-level estimates were quite stable, most of the individual-level estimates were not. The only parameters that had 100% of their $\hat{R}'s$ below the acceptable level of 1.1 (where $\hat{R}'s$ closer to 1 are best) were the boundary (response cautiousness) and drift (evidence accumulation speed) parameters, hence their focus in the main text. The parameters used to calculate the bias (Abeta and Bbeta) had the majority of individual estimates as acceptable, but a good size of them were not. It is for this reason that we did not feel comfortable speculating on bias across groups. However, the most shocking case was in the individual estimates of the non-decision time: only a bit over 4% of the estimates were deemed acceptable, with the highest $\hat{R}$ being about 2.6. The reader should consider that the ideal Gelman-Rubin is 1, and only a deviation of 0.1 is generally considered tolerable, which should put into perspective how terrible of a convergence an $\hat{R}$ of 2.6 really is. This almost certainly comes from the large amounts of missing data from both clinical groups. Additionally, these Gelman-Rubin results are quite general across the models: while the exact percentages of acceptable $\hat{R}'s$ obviously differ across models, the general pattern is the same. In fact, the pattern is a bit worse in the regression models, since they are using the same amount of data to estimate more parameters, which seemed to make the estimation of all parameters suffer. This latter occurrence is a large reason why we did not focus on the regression models in the main text.

These issues are important to highlight, because a poor $\hat{R}$ is reflective of the estimation procedure giving extremely inaccurate estimates of the true posteriors. This inaccuracy, then,

leads to incorrect and unjustified assertions about the differences between groups. The Ratcliff Diffusion Model, for all its accomplishments, is still a very complex model, a glutton for data, and unfortunately it would seem that the ambulatory setting of the current paper is too much of a strain on the model for certain of its parameters. As mentioned in the main text, this is a great shame, because of the known associations between HD and motor control, and the intended relationship between the NDT and motor control in the diffusion model (see the main text for references). This fact should then highlight the significance of the main findings of the paper: that even with all the complications that come with a dataset from an in-home assessment, the diffusion model is just powerful enough to detect an otherwise elusive result with at least one of its parameters.

**More on Individual Differences**

In the main text, we mentioned that the main model that was used, the variability model in this supplement, included a measure of individual difference spread in the groups that we simply labeled *variability*. In this section, we elaborate a bit more on this concept, as we believe it may have important ramifications for any practical implementation of this model. Indeed, that it was the best non-regression model, and that the better fitting regression models also allowed for changes in variability between groups, should warrant further inquiry. Figure 7 illustrates an artificial example.

The means and standard deviations for each estimate are reported in Table 3 of the main text. While we did not find any credible differences using the HDI criterion, we wish to highlight some results that were noteworthy. We found that there was a 93% chance that controls > HD in variability, and that there was only a 56% chance that HD < pre-manifest in variability. This latter result is particularly interesting, as the symmetry in the difference distribution (HD - pre-manifest) indicates that the mean difference in variability between HD and pre-manifest was almost zero. This suggests that both clinical groups may have a similar spread of individual scores, but one is shifted slightly to the right compared to the other. (This can be seen in Figure 5, for instance.) As for across-conditions, within-group comparisons, the only notable comparison was that there was a 79% chance that Few < Many variability with controls only. These relatively high probabilities suggest that, with more data, this issue of the spread of individual differences across the groups may be worth further testing and scrutiny.

Now, with those numbers under our belts, we should know how to use them. First, we should understand the problem as tackled in the main paper, compared to how a clinician might use the model. In the main text, we were given the group assignments of each person (full Huntington's,

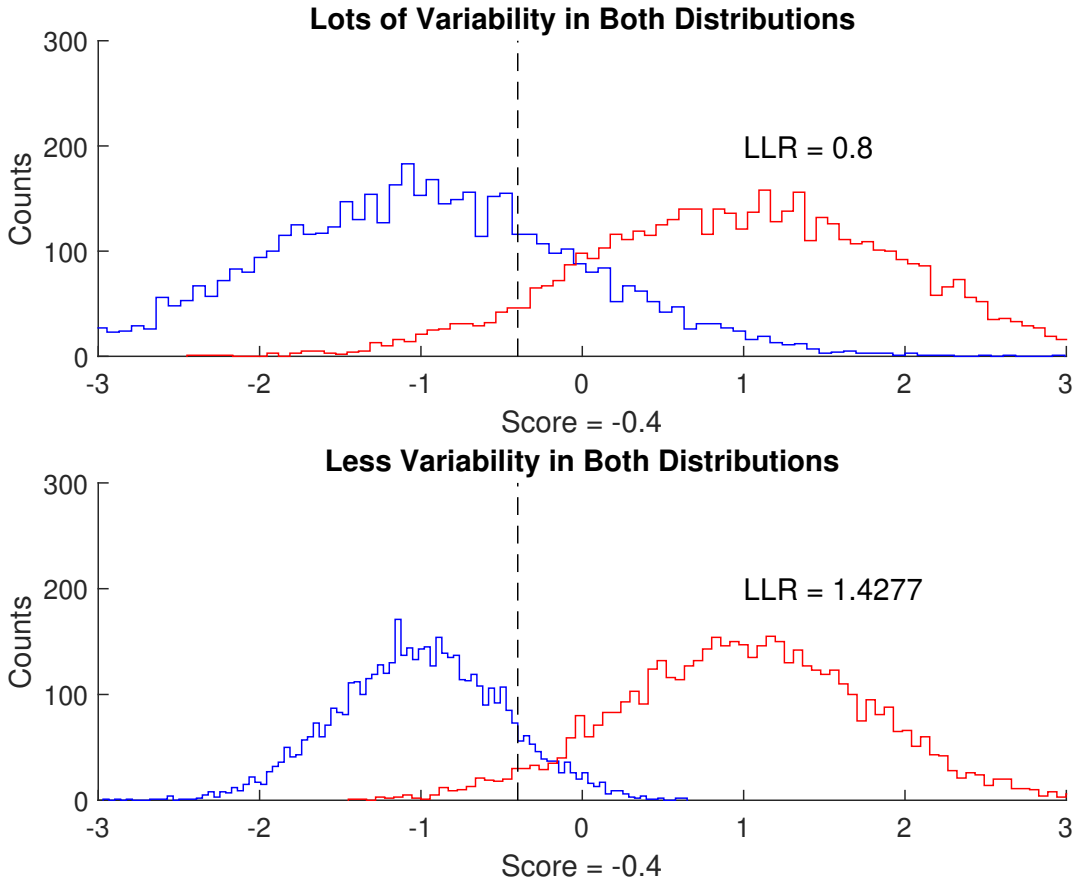



Figure 7: An example showing the importance of the spread of individual scores within each group. With narrower individual spread within a group, we can be more confident in our group assignments once we have an observed score, which can potentially be immensely helpful in a predictive scenario in a clinical setting. The "LLR" is the log-likelihood ratio of choosing the blue over the red distribution when deciding from which distribution the observed score of -0.4 came. Larger LLRs indicate more evidence that the score came from the blue distribution. Having a smaller spread nearly doubled the log-likelihood. The particular parameters that were used to simulate this example were: (Top) Blue mean = -1, blue standard deviation = 1; red mean = 1, red standard deviation = 1; (Bottom) blue mean and standard deviation = (-1, .5); red mean and standard deviation = (1, .75).

pre-manifest, control), and the model set out to see, *given these group assignments*, what differences were observable between the groups. The central finding of the main text was that the groups differed in their drift rates. Now, the clinician who may find this result useful needs to consider that s/he has the opposite problem: the clinician is given a score, and from this must determine the appropriate group assignment for their patient. The posterior distributions of mean drift rate within each group, found using the data from the main text, can assist in this group assignment, as these distributions give the clinician some prior information about what to expect from each group in terms of drift rate scores. That is, the clinician is faced with a signal detection problem (Green, Swets, et al., 1966), and the results of this paper give the distributions that s/he can use to decide what is "signal" (e.g., pre-manifest) and what is "noise" (e.g., healthy control). This may be particularly helpful in the service of tracking disease progression, or even progress within a treatment regimen.

Now, to elaborate further, consider the example given in Figure 7 of this supplement. (The numbers are artificial and do not correspond to what was actually found in the main text, but the reader can hopefully excuse the simplification in the service of clarity.) Suppose that the clinician found a score of -0.4 for a patient, and must decide if the patient should be classified as, say, pre-manifest (red) or fully HD (blue). The top portion of Figure 7 shows a common and sometimes frustrating situation in clinical modeling: that of great variability in scores within each group. What the clinician can do in this case, say, is take the approximate density at the observed score for both the blue and red distributions, and take a ratio of the blue density to red density. This gives the likelihood that the patient came from the blue versus the red distribution of scores. The likelihood numbers are typically too small to be meaningfully understood, so taking a log is done to make the numbers more understandable. Doing so produces the "LLR" (log-likelihood ratio) given in the top portion of Figure 5. The clinician might come up with a simple rule, such as, "If the LLR is above zero, there is more evidence that the patient came from the blue distribution [HD group]." However, with only an LLR of 0.8, the clinician may not be particularly confident in the assignment.

The situation is a bit better in the bottom portion of Figure 5, where the spread of each distribution has been reduced (but the means remain the same). Now, using the same score, but with these different distribution statistics, the LLR is seen to be about 1.42, almost double what was observed in the previous situation. While not astronomically different, the point to be made here is that *less spread of individual scores in a group means that an observed score close to the mean of*

*the group will give the clinician greater confidence in assigning the new, unknown patient to that lower-variability group.*

We imagine that this classification scheme may be particularly useful to clinicians who may want to monitor patient performance over a sequence of testing sessions. For instance, a researcher may want to chart the estimated individual drift for the person, and at the same time look at the log-likelihood ratio of that person being in one group versus another (pre-manifest vs control, or HD vs pre-manifest, for instance). This may allow the researcher to have a set of scores that tracks progress while the patient undergoes some experimental treatment regimen, for instance. A substantial change in the log-likelihood of being in the severe patient group may be evidence, then, that the treatment regimen is working as intended. This kind of classification-based progress reporting could be done using the posterior distributions that we observed in the main text. And, since we found some preliminary evidence that the HD group had less individual difference than the other groups[4], a practical implementation of the current work would lead to something akin to the bottom panel of Figure 7, with more confidence in the assignment of group membership.

In particular, an interested researcher could use the current model in a future application of the current experimental paradigm, and update the prior distributions on the group mean drift rates using the estimates of $\mu$ and $\sigma$ (group drift and variability, resp.) as the prior estimates of group mean and group spread of scores in a subsequent fitting of the model.

## References


Green, D. M., Swets, J. A., et al. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Academic Press.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2010). Hierarchical diffusion models for two-choice response times. *Psychological methods*, *16*(1), 44.


[4]While we did not find evidence of differences in variability across groups in either the non-regression or regression models, it may be the case, given the relatively high probabilities in some cases, that more data may shed more light on this issue.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*(5), 1413–1432.

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... others (2018). Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, *25*(1), 35–57.

Walker, F. O. (2007). Huntington's disease. *The Lancet*, *369*(9557), 218–228.