School of Psychology

A Bayesian Evaluation of Subtyping Methods in Parkinson's Disease

Andrew Robert Johnson

This thesis is presented for Degree of

Doctor of Philosophy

of

Curtin University

October 2019

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research (2007) –updated March 2014. The data analysed were collected by a research study which received human research ethics approval from the Curtin University Human Research Ethics Committee (EC00262), Approval Number #HR158/2013

Signature: ……

Date: ……

Acknowledgements

Thanks first must go to my wonderful supervisors: Prof Natalie Gasson, A/Prof Andrea Loftus, and Prof Romola Bucks. You all started me on this journey and have played no small part in inspiring my passion and curiosity for research. Without your encouragement to learn and challenge myself (and when to stop learning and start writing) I would not have come this far and achieved this much.

To my friends, your patience and acceptance of my inability to learn the meaning of 'work/life balance' will always be appreciated. You all made things that little bit easier for me, especially when I wasn't making things all that easy for myself. A special thanks must go to Jack, Severine, and Stephen, for experiencing the highs and lows of my journey every week on the golf course.

To my partner, Jordina, you've been eternally patient throughout my journey. Through the late nights, the stressful days, and the sleep-ins while you were running; you were there. I will always be grateful to have had such an understanding and passionate cheerleader in my corner. I can't wait to see what our life together looks like without my thesis as a third partner!

Finally, to my parents, I owe the most of all. Your unconditional support and encouragement have given me the courage to take on challenges that I never would have thought possible. You raised me with the mindset that there was nothing that I couldn't do, as long as I worked hard enough. Now look where I am! Thank you for supporting me, for guiding me, and for always inspiring me.

Table of Contents

List of Figures

List of Tables

Abstract

Parkinson's disease (PD) is gaining recognition as having significant heterogeneity in its clinical presentation. To explain this heterogeneity, several motor subtypes of the disease have been proposed. These subtypes can be separated into two main approaches: empirical and data-driven. Empirical approaches are characterised by the researcher choosing which symptoms, and the severity cut-offs for symptoms, to use in grouping individuals. In contrast, data-driven approaches use statistical methods to determine the number of groupings (if any) and the appropriate measures and cut-offs that can be used to separate them. These subtyping approaches make assumptions about the ways in which different symptoms change over time, the ability to measure different symptoms, and the relationships between different symptoms within a given disease subtype. While these assumptions may be appropriate for PD subtyping, current statistical approaches are not able to test whether this is the case. The ability to test these assumptions is found in the field of Bayesian statistics. The present research aimed to evaluate the assumptions underlying current approaches to PD subtyping and subsequently develop a new model of motor subtypes in PD. Four studies were conducted.

Study 1 presented the first analysis of the progression of tremor and postural symptom severity in PD. The aim of Study 1 was to evaluate the assumption underlying the most common empirical approach to PD motor subtypes: Jankovic et. al.'s (1990) tremor-dominant (TD) and postural instability and gait difficulty (PIGD) subtypes. The TD and PIGD subtypes are based on the relative difference in severity between tremor and postural symptoms. However, this approach assumes that the difference between symptoms is (approximately) constant over time. Study 1 identified that tremor symptoms showed no significant change in severity over time (β = 0.004 [-0.03, 0.04], p = .831). The analyses further identified that postural symptoms showed a significant yearly increase in severity (β = 0.11 [0.09, 0.13], p <

.001). Initial severity and rate of change in tremor symptoms were not related to the initial severity or rate of change in postural symptoms. Study 1 concluded that this faster progression of postural symptom severity meant that the difference in severity between tremor and postural symptoms would not be constant over time. Consequently, the TD and PIGD subtypes would likely be heavily influenced by the disease duration of the individual.

Study 2 presented the first evaluation of PD motor symptom measurement using Bayesian structural equation modelling (BSEM). A concern with current subtyping approaches is the lack of methodological validation for approaches to measuring motor symptoms. Consequently, it is difficult to identify the measurement quality of symptoms being used for subtyping. Study 2 identified six primary motor symptoms: axial, resting tremor, postural and kinetic tremor, rigidity, akinesia, and lower-body akinesia. Study 2 further identified that the severity of these six motor symptoms differed by laterality (i.e., left- and right-sided severity). This model of motor symptoms was also replicated in a separate sample of de novo individuals with PD. The key finding of Study 2 was that the ratings on a given item (designed to assess a particular symptom; e.g., rigidity), were being influenced by the severity of other motor symptoms (e.g., akinesia). This complexity in measurement could not be captured using sum or mean scores. Study 2 concluded that the previous subtyping analyses using sum or mean scores to represent motor symptom severity could be biased by this approach to measurement. Study 2 further developed a spreadsheet that clinicians could use to derive motor symptom severity ratings (including laterality of severity) for a single individual.

Study 3 developed a model of PD subtypes using a Bayesian extension of latent profile analysis. Study 3 sought to develop and propose a method of PD subtyping that better represented the clinical presentations of PD heterogeneity as well as being more statistically powerful than current methods. The method is the first of its kind to allow for the measured

symptoms within each subtype to be related and is the first that can simultaneously account for measurement error, antiparkinsonian medication, missing data, and correctly handle categorical data. The analyses identified that three motor subtypes were present. The "Postural Kineto-Rigid" (PKR) subtype was defined by increased severity of postural, akinetic/bradykinetic, and rigidity symptoms. The "Tremulous General Severity" (TGS) subtype was defined by increased overall severity and markedly increased resting tremor severity. The "Asymmetric Undifferentiated" (AU) subtype was defined by a mixture of the symptoms of the first two subtypes, but with a negative correlation between the severity of symptoms of either laterality. In other words, an increased level of severity on one side of the body was associated with a decreased level of severity on the other. Current approaches to PD subtyping assume that these correlations within subtypes are not present, which Study 3 demonstrated is not the case. Consequently, Study 3 concluded that the assumptions of current data-driven approaches are not appropriate and should be circumvented using Bayesian methods.

Study 4 used path analysis to demonstrate a clinically relevant difference in quality of life between the people of each of the motor subtypes that were derived in Study 3. Individuals belonging to the PKR type experienced significantly poorer quality of life across several domains than individuals in the AU type. The largest differences were seen in activities of daily living, mobility, and cognition. These significant differences, and their presence only in the PKR type, indicates a clear clinical application for PD motor subtyping. The ability to identify individuals at greater risk of poor quality of life in the future through a simple motor examination could be of prognostic value in clinical practice. Thus, Study 4 concluded that the present results indicate a proof-of-concept for the potential of motor subtyping as a tool for clinical practice.

Overall, the present thesis identified that the assumptions of current approaches to motor symptom measurement and subtyping are not consistent with the clinical presentation of Parkinson's disease. These restrictive assumptions can be avoided by taking advantage of the flexibility available in Bayesian statistics. A further contribution of the present thesis was an accessible introduction to applying and interpreting Bayesian statistics in PD. While this area of statistics has a large range of potential applications in PD, it also has a high barrier to entry for researchers and clinicians. The present thesis aimed to reduce this barrier by providing a practical example of the benefits of the approach while still emphasising the clinical relevance of the findings.

Chapter 1 - General Introduction

1.1 Parkinson's Disease

The first references to Parkinson's disease (PD) date as far back as 1000BC India, but it was first brought to the attention of medicine through the now infamous 'An Essay on the Shaking Palsy' by James Parkinson in 1817 (Goetz, 2011). Parkinson described the disease presentation as being primarily characterised by "involuntary tremulous motion" (Goetz, 2011). However, the present day understanding of PD symptomology has expanded to cover almost all domains of human functioning (Politis et al., 2010). PD symptoms can be organised into four primary categories: motor, autonomic, cognitive, and psychiatric (Politis et al., 2010).

1.2 Motor Symptoms

1.2.1 Tremor

There are four cardinal motor symptoms in PD: tremor, rigidity, bradykinesia/akinesia, and postural instability (Postuma et al., 2015). Perhaps the most widely-recognised motor symptom of the disease, Parkinsonian tremor can be separated into three categories: resting, postural, and kinetic (Goetz et al., 2008). Resting tremor refers to a tremor that is present when the individual is not moving (i.e., at rest). One of the most common resting tremors is the so-called 'pill-rolling' tremor, which presents as small, circular motions of the thumb over the top of the forefingers (Sveinbjornsdottir, 2016). Postural tremor refers to a tremor that is present when the individual is maintaining a posture of some form (Sveinbjornsdottir, 2016). One of the most common assessments of this type of tremor involves an individual extending their arms in front of them and maintaining that pose (Goetz et al., 2008). Kinetic tremor, as the name suggests, refers to a tremor that is present during a movement itself (Sveinbjornsdottir, 2016). The common assessment for this type of

tremor involves asking an individual to touch their nose and then touch the assessor's finger (Goetz et al., 2008).

1.2.2 Rigidity

Rigidity in PD can present in two ways: "leadpipe" and "cogwheel" (Broussolle et al., 2007). These two forms are differentiated by the uniformity of the rigidity throughout movement (Broussolle et al., 2007). The "leadpipe" form presents as a uniform level of rigidity at all ranges of movement of the joint (Broussolle et al., 2007). In contrast, the "cogwheel" form presents as regular interruptions of the movement of the joint (similar to non-constant movement of a cogwheel; Ghiglione, Mutani, & Chio, 2005).

1.2.3 Bradykinesia/Akinesia

Bradykinesia and akinesia refer to the amplitude of produced movement (or lack thereof; Jankovic, 2008). Bradykinesia describes a reduction in the amplitude of produced movements (Jankovic, 2008). Some common presentations of bradykinesia include a reduced size of handwriting, expressive gestures, and even reduced speech volume (Politis et al., 2010). Where bradykinesia describes a reduced amplitude of movement, akinesia instead describes a reduced *production* of movement (Jankovic, 2008). One the of most common presentations of akinesia is facial masking, where individuals lose the ability to produce facial expressions without explicit initiation (Tickle-Degnen & Doyle Lyons, 2004). Facial masking presents as a constant 'blank' facial expression, regardless of the individual's emotional reaction to a given situation (Tickle-Degnen & Doyle Lyons, 2004)

1.2.4 Postural instability

The last cardinal motor symptom of PD is postural instability, or the inability to maintain and correct one's balance (Sveinbjornsdottir, 2016). As a consequence of postural instability individuals with PD are at a greater risk of falls and associated health

complications (Morris, Iansek, Smithson, & Huxham, 2000). Additionally, postural instability is often the least responsive of the motor symptoms to antiparkinsonian medication (Morris et al., 2000). Generally, these symptoms need to be managed through movement training and attentional strategies, rather than pharmacological intervention (Morris et al., 2000).

1.3 Autonomic Symptoms

PD presents with autonomic dysfunction across a range of bodily domains (Visser, Marinus, Stiggelbout, & Van Hilten, 2004). Gastrointestinal dysfunctions are a common experience (Politis et al., 2010). Individuals can report difficulties with constipation, loss of appetite and weight loss, as well as an overproduction of saliva (Micieli, Tosi, Marcheselli, & Cavallini, 2003). Sleep dysfunctions are increasingly acknowledged as symptoms associated with PD (Pushpanathan, Loftus, Thomas, Gasson, & Bucks, 2016). Common disturbances include REM sleep behaviour disorder (RBD), insomnia, and parasomnia (Pushpanathan et al., 2016). Urinary dysfunction is also common, with individuals experiencing difficulties with incontinence, nocturnal urination, and frequency (Micieli et al., 2003).

1.4 Cognitive Symptoms

The extent of cognitive impairments in PD is a complex topic in and of itself. Individuals with PD can present with deficits in both frontal (e.g., executive functions) and global cognitive abilities (Litvan et al., 2012). There is also a higher prevalence of dementia in PD, compared to the general population (Aarsland, Andersen, Larsen, Lolk, & Kragh-Sorensen, 2003). There has been some discussion as to whether the presence of frontal cognitive deficits are indicative of an eventual onset of dementia (Foltynie, Brayne, Robbins, & Barker, 2004). However, this is still a topic of much debate (Litvan et al., 2012).

1.5 Psychiatric Symptoms

Psychiatric complications are a common comorbidity of the disease, and have a significant impact on the quality of life of the individual (Weintraub, Moberg, Duda, Katz, & Stern, 2004). One of the most common comorbidities is depression, with an estimated prevalence of up to 50% for any type of depressive disorder (Weintraub & Stern, 2005). The presence of a depressive disorder in PD is associated with significantly poorer clinical outcomes for the individual and higher levels of caregiver stress (Weintraub & Stern, 2005). A less common, but similarly disabling, complication is psychosis (Weintraub & Burn, 2011). Individuals with psychosis will most commonly experience visual hallucinations, however hallucinations within the other sensory domains have also been seen (Weintraub & Burn, 2011). A psychiatric symptom of PD that has only risen to prominence in the past decade is that of impulse control disorders (ICDs; Weintraub et al., 2010). These disorders present as an inability to regulate particular behaviours (e.g., gambling, shopping, or sex; Weintraub et al., 2010). Given that some of these behaviours may be embarrassing for the individual to discuss with their practitioner, ICDs are considered to be an underreported symptom of PD (Weintraub & Burn, 2011).

It is clear that PD symptomology is extensive and present in almost all aspects of an individual's ability to function. This breadth of affected areas can likely be attributed to the widespread neurological degeneration that occurs as part of the disease (Halliday & McCann, 2010).

1.6 Parkinsonian Degeneration

1.6.1 Dopaminergic

The hallmark degeneration associated with PD is the loss of dopaminergic neurons in the pars compacta of the substantia nigra (Davie, 2008). The substantia nigra is a midbrain

structure with two components: the pars reticular (SNr) and the pars compacta (SNc; Johns, 2014). The substantia nigra is heavily involved in the modulation of voluntary movements and reward behaviours (Hodaie, Neimat, & Lozano, 2007). This modulation is primarily achieved through the projections of dopaminergic neurons from the SNc to the caudate nucleus and putamen within the striatum (Hodaie et al., 2007). These dopaminergic neurons see a heavy degeneration in PD, with approximately 89% of caudate neurons and 98% of putamen lost at the time of death (Rajput et al., 2008). The nigrostriatal degeneration of dopaminergic neurons was the first neuropathology proposed to underlie the disease (Brissaud, 1925). However, with advancements in neuroimaging and histopathology came a deeper understanding of the other neurological systems affected by the disease.

1.6.2 Cholinergic

A secondary neurotransmitter implicated in PD symptomology is acetylcholine (Bohnen & Albin, 2011). There are two primary sources of cholinergic neurons in the brain, both of which degenerate in PD: the basal forebrain complex and the pedunculopontine nucleus (PPN; Bohnen & Albin, 2011). Within the basal forebrain complex the key cholinergic structure is the nucleus basalis of Meynert (nbM), which supplies the majority of the cortex (Yarnall, Rochester, & Burn, 2011). The PPN projects to, and receives projections from, the basal ganglia (Yarnall et al., 2011). Degeneration of cholinergic neurons within these structures is frequently seen in PD (Bohnen & Albin, 2011). These structures are, however, also significantly affected by Lewy pathology (Halliday & McCann, 2010).

1.6.3 Lewy Pathology

A final key component of Parkinsonian degeneration is the presence of Lewy pathology. Lewy pathology refers to the presence of two types of neuronal inclusions: Lewy bodies and Lewy neurites (Braak et al., 2003). Lewy bodies are aggregates of misfolded proteins found in nerve cells (Olanow & Brundin, 2013). These neuronal inclusions typically

impinge on neighbouring neurons and other cellular components, impeding their functioning and eventually causing cell death (Olanow & Brundin, 2013). Lewy neurites are a similar, but distinct, pathology where the accumulations are present in the neurites themselves (Olanow & Brundin, 2013). In PD, both Lewy bodies and Lewy neurites are accumulates of the protein *α*-synuclein (Spillantini & Goedert, 2000). Lewy pathology in PD is present in the majority of neurological structures previously described, and becomes progressively more severe (Braak et al., 2003).

1.6.4 Progression of Degeneration

A key characteristic of Parkinsonian neurodegeneration is its progressive nature (Davie, 2008). Braak and colleagues proposed the most widely used characterisation of PD's progressive degeneration in 2003. Braak et al. (2003) proposed six 'stages' of ascending neurodegeneration in PD, progressing from the lower brainstem to the neocortex. According to Braak et al. (2003), the disease begins in the medulla oblongata and anterior olfactory nucleus before progressing to the locus ceruleus (Stages 1-2). The disease then begins to present in the substantia nigra and nbM and subsequently progresses to the amygdala and thalamus (Stages 2-4; Braak et al., 2003). Finally, the disease progresses to the neocortex (Stages 5-6; Braak et al., 2003). These Braak stages were intended to provide a consistent neuropathological 'roadmap' to the degeneration and associated clinical manifestations of PD (Braak et al., 2003). However, despite their wide acceptance and use, subsequent evaluations of these have shown that both symptom presentation and neurodegeneration in PD is much more heterogeneous than can be captured by this staging hypothesis (Jellinger, 2009). The heterogeneity in both disease presentation and progression has been the subject of much debate in the PD literature (Jankovic et al., 1990).

1.7 Heterogeneity in Symptom Presentation

The heterogeneous clinical presentation of PD is a well-recognised, but little understood, facet of the disease (Kehagia, Barker, & Robbins, 2010). While there is the previously mentioned individual variation in neurodegeneration and progression, there is also marked individual variation in symptomatic presentation (Jellinger, 2009).

1.7.1 The Introduction of PD Motor Subtyping

A cornerstone study in PD heterogeneity was conducted by Zetusky, Jankovic, and Pirozzolo (1985). The authors assessed differential relationships between the cardinal motor symptoms and age at onset, functional disability, and mental status (among others; Zetusky et al., 1985). Zetusky et al. (1985) identified that increased severity of postural instability symptoms was associated with poorer mental status, an older age at onset, and increased functional disability. These relationships were not present with increased severity of tremor symptoms (Zetusky et al., 1985). Based on these differences in clinical outcomes, Zetusky et al. (1985) proposed that there were two 'subgroups' of PD: individuals with dominant postural instability symptoms, and individuals with dominant tremor symptoms.

This conclusion, however, was limited by the authors' approach to measuring motor symptoms. Zetusky et al. (1985) used a rating system that was developed by the second author in 1981 (Jankovic & Frost, 1981). Jankovic and Frost's (1981) scale was developed using a sample of only 11 individuals, and had yet to be validated in another sample prior to the study by Zetusky et al. (1985). It is, therefore, unclear how accurate the subtyping results of Zetusky et al. (1985) are, given the unknown validity and reliability of the measurement system used. This was a common difficulty for studies of that time, as there was little consistency between studies in the rating scales used for measuring motor symptoms (Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease, 2003).

1.7.2 The Unified Parkinson's Disease Rating Scale (UPDRS; 1987)

The publication of the Unified Parkinson's Disease Rating Scale (UPDRS) in 1987 brought some much-needed standardisation to the measurement of Parkinsonian symptoms (Fahn, Elton, & UPDRS program members, 1987). The UPDRS introduced standardised, validated rating scales for the measurement of both motor and non-motor PD symptoms (Fahn et al., 1987). These UPDRS scales soon became the most widely used approach to symptom measurement in PD (Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease, 2003). Of particular relevance to PD motor heterogeneity was the motor examination (Part 3 of the UPDRS; Fahn et al., 1987). The introduction of the UPDRS motor examination allowed for the replicable quantification of motor symptom severity, which in turn allowed for a replicable means of grouping individuals.

1.7.3 Popularising Motor Subtyping

The 1987 UPDRS was a key component of what became the most influential study of motor heterogeneity in PD (Jankovic et al., 1990). Jankovic et al. (1990) used the UPDRS to propose a means of formally classifying the subgroups initially discussed in Zetusky et al. (1985). The authors selected several items from the UPDRS and created two subscales, one assessing tremor symptoms and one assessing postural and gait difficulties (Jankovic et al., 1990). Based upon the ratio of these subscale scores, Jankovic et al. (1990) separated individuals into two subgroups: one group with predominate severity of tremor symptoms, and one group with predominate severity of postural symptoms. Jankovic et al. (1990) denoted these the “Tremor Dominant” (TD) and “Postural Instability and Gait Difficulty” (PIGD) subtypes, respectively. The authors identified poorer cognitive performance and greater motor impairments in the PIGD subtypes, and these findings have since been replicated (Jankovic et al., 1990; Thenganatt & Jankovic, 2014). Jankovic et al.’s (1990)

study was highly influential and further investigation into means of grouping individuals by motor symptom severity soon followed.

1.7.4 Extending Motor Subtyping to Akinesia and Rigidity

Where Jankovic and colleagues (1990) used the UPDRS to quantify subtypes defined by postural symptom severity, Schiess, Zheng, Soukup, Bonnen, and Nauta (2000), instead, defined subtypes according to akinesia and rigidity symptoms. A presentation of PD dominated by the severity of akinesia and rigidity symptoms had been discussed several decades prior by Hoehn and Yahr (1967). However, as with the initial discussion of the PIGD subtype by Zetusky et al. (1985), the lack of a standardised form of motor symptom measurement limited its exploration and replication in the wider literature. Schiess et al. (2000) used the, now well-established, UPDRS to derive a means of classifying the so-called "akinetic-rigid" (AR) subtype. However, Schiess et al.'s (2000) classifications did not reach nearly the same level of use as those of Jankovic et al. (1990).

At the time of writing, Web of Science reports that Schiess et al.'s (2000) study garnered 71 citations, whereas Jankovic et al. (1990) has received 698. This was likely due to several reasons. Firstly, Schiess et al. (2000) developed their classifications in a sample of only 30 individuals, calling into question the generalisability of their approach. Further, while the authors demonstrated pathological differences between TD and AR individuals, they did not investigate differences between PIGD and AR individuals (Schiess et al., 2000). It was not clear whether the AR subtype was truly distinct from the PIGD subtype, or simply another means of measuring the same construct. While the work of Schiess et al. (2000) was not as influential as that of Jankovic et al. (1990) it is still an important milestone in the development of the AR subtype.

1.7.5 Emergence of Data-Driven Methods

It was around this time that the first demarcation of subtyping methods into "empirical" and "data-driven" approaches emerged (Graham & Sagar, 1999). Under the empirical approach the subtypes are known *a priori* and the researchers are developing a means of measuring them (van Rooden et al., 2010). The subgroups presented by Jankovic et al. (1990) and Schiess et al. (2000) are prototypical examples of the empirical approach. In contrast, the data-driven approach is exploratory (van Rooden et al., 2010). Data-driven approaches to subtyping rely on statistical analyses to identify groups (or 'clusters') in a given set of symptom measurements (van Rooden et al., 2011).

Data-driven subtyping was first introduced to the PD literature by Graham and Sagar (1999). The authors criticised the empirical approach for relying solely on researcher judgement, rather than supplementing with statistical methods, in the choice of defining symptoms for each subtype (Graham & Sagar, 1999). The authors used a *K*-means cluster analytic approach which seeks to identify *K* different groups whose mean are as different as possible from each other (Jain, 2010). The authors applied this cluster analytic approach to several measures of motor function, affect, and cognition to identify five groupings in their sample of 176 individuals with PD (Graham & Sagar, 1999). The first group was denoted "motor only", defined by higher motor symptom severity with no difference in cognitive performance (Graham & Sagar, 1999). The second group was denoted "motor and cognitive", characterised by both increased motor severity and reductions in executive functioning (Graham & Sagar, 1999). The third group was denoted "rapid progression", characterised by an older age at onset of the disease and a faster progression of degeneration compared to the others (Graham & Sagar, 1999). The authors did not name the final two groups but described them as being characterised by an increased disease duration (Graham & Sagar, 1999).

Graham and Sagar's (1999) work was received in a manner similar to that of Schiess et al. (2000). While it was the first study in what has become an almost saturated area (data-driven subtyping), it gained little recognition (Web of Science lists the citation count at 93). It would be another five years before data-driven subtyping would make another, more successful, appearance.

1.7.6 Popularisation of Data-Driven Subtyping Methods

The work of Lewis and colleagues (2005) represents another pivotal moment in the PD heterogeneity literature. Lewis et al. (2005) shared Graham and Sagar's (1999) perspective on the benefits of a cluster analytic approach but felt that their approach had limitations that required addressing. Lewis et al. (2005) presented two, key criticisms of Graham and Sagar (1999). First, they argued that individuals with advanced Parkinsonism should not have been included in the analysis (Lewis et al., 2005). According to Lewis et al. (2005), those in the advanced stages of the disease would not present with as much clinical variability as those in early stages of the disease. Second, they argued that by including all available measures in the subtyping analysis, Graham and Sagar (1999) were left with no means of validating their derived subtypes. Despite the importance of this second criticism, this would become a common approach in subsequent research in PD subtyping.

Lewis et al. (2005) included measures of mood, quality of life, cognition, and Parkinsonian severity (disease duration, motor severity, and dopaminergic medication). The authors identified four subtypes: younger onset, tremor dominant, non-tremor dominant, and rapid motor progression (Lewis et al., 2005). However, their work was also not without limitation. Whilst the measures of mood, cognition, and quality of life were all established and validated scales, the authors included a number of self-derived measures of Parkinsonian severity (Lewis et al., 2005). Their approach to quantifying the rate of disease progression was to divide the sum of UPDRS sections I-III by the number of years since the individual's

diagnosis of PD (Lewis et al., 2005). The result of this approach was a rough analogue of the amount of disease severity per year of disease experience. This estimate may be not be accurate for two reasons.

Firstly, the time of disease diagnosis is a poor proxy for time of disease onset (Tolosa, Wenning, & Poewe, 2006). This is because, not only can disease onset precede diagnosis by several years, but the symptoms with which individuals present can influence the time it takes for a diagnosis to occur. For example, individuals with PD can note a loss of smell several years before their PD diagnosis (Tolosa et al., 2006). An individual with a resting tremor is more likely to seek medical advice, and receive a PD diagnosis sooner, than an individual with a loss of smell.

Of more concern is the authors' quantification of "tremor" and "non-tremor" symptoms. For these two motor symptoms, Lewis et al. (2005) selected several items from the UPDRS motor assessment and calculated averages for each individual. This was the same approach used by Jankovic et al. (1990) to quantify tremor and postural severity, and shares the same limitations. While the selected UPDRS items may appear to measure the motor symptoms of interest, this was not statistically or methodologically confirmed by either study (Jankovic et al., 1990; Lewis et al., 2005). Without assessing the measurement properties of these author-proposed scales, it cannot be said with confidence that the motor symptoms of interest are being adequately represented. This lack of established subscales for individual motor symptoms (e.g., rigidity or akinesia) would continue to limit the PD subtyping studies that followed.

1.7.7 The Introduction of Latent Profile Analysis

Following the work of Lewis et al. (2005), k-means cluster analysis saw increasing use in PD (P. Liu, Feng, Wang, Zhang, & Chen, 2011; Merchant, Luciana, Hooper, Majestic,

& Tuite, 2008; Reijnders, Ehrt, Lousberg, Aarsland, & Leentjens, 2009; Schrag, Quinn, & Ben-Shlomo, 2006). However, some researchers began to question the suitability of the analysis itself (van Rooden et al., 2010). While popular, the statistical assumptions and mechanics underlying k-means limit its accuracy when compared with other approaches (Magidson & Vermunt, 2002b). van Rooden et al. (2011) argued that a more appropriate analysis for PD was latent profile analysis (LPA; also referred to as "model-based clustering"). LPA is a much more flexible approach to clustering, which better handles outliers and unequal cluster sizes (Vermunt & Magidson, 2002). By using LPA in place of k-means for the subtyping analysis, van Rooden et al. (2011) aimed to achieve a more precise 'picture' of the nature of heterogeneity in PD.

van Rooden et al.'s (2011) cluster analysis included measures of a range of clinical features; including motor severity, autonomic dysfunction, and sleep problems (among others). The study included a sample of 415 individuals with idiopathic PD and four clusters were identified (van Rooden et al., 2011). The first cluster was characterised by mild severity across all measures, the second by severity on motor measures, the third by severity on nondopaminergic symptoms, and the fourth cluster by increased severity on all symptoms (van Rooden et al., 2011). While van Rooden et al.'s (2011) use of LPA represents a significant step forward in PD subtyping methods their application was limited by several methodological and statistical issues.

Statistically, van Rooden et al.'s (2011) application of LPA introduced a number of problems. While LPA, in and of itself, is exceedingly flexible and powerful, it depends on a statistical assumption that is not consistent with the aim of van Rooden et al.'s (2011) study. The key statistical assumption of LPA is the assumption of local independence (Vermunt & Magidson, 2002). Under the assumption of local independence, within each of the identified clusters (or subtypes) there should be no relationship between any of the subtyping variables

in the analysis (Vermunt & Magidson, 2002). In other words, the presence of subtypes should explain all observed covariation between variables (Magidson & Vermunt, 2002a). Given that all individuals in the analysis have the same underlying disease (PD), it is likely that some observed relationships will be a consequence of the disease (and universal to all individuals) rather than a consequence of a specific subtype. A fuller introduction to k-means and LPA, as well as discussion of the assumption of local independence, and its implications for subtyping accuracy, is provided in Chapter 5. This flaw, however, is yet to be addressed in the PD subtyping literature.

1.8 Current State of PD Subtyping

The PD subtyping field is currently divided between two primary schools of thought: studies that continue to apply Jankovic et al.'s (1990) TD and PIGD classifications, and studies that apply LPA to various combinations of clinical outcomes. Despite the advancements in the field, and in the understanding of Parkinsonian heterogeneity, Jankovic et al.'s (1990) TD and PIGD subtypes still see wide use in new research (e.g., Beretta, Vitorio, Santos, Orcioli-Silva, & Gobbi, 2019; J. B. Liu et al., 2019; Pinter et al., 2019). While this approach to subtyping has received criticism in the past (Graham & Sagar, 1999), Jankovic et al.'s (1990) TD and PIGD classifications have yet to be convincingly demonstrated as flawed or unsuitable. It is likely that the TD and PIGD subtypes will continue to be used until such a demonstration is provided.

A similar concern exists with the increasing number of PD subtyping studies using LPA. Following the publication of van Rooden et al.'s (2011) approach to subtyping, LPA has been used in PD with measures of cognition (LaBelle, Walsh, & Banks, 2017), depression (Starkstein et al., 2011), and psychosis (Fereshtehnejad et al., 2015). However, there has yet to be any discussion of the impact of the local independence assumption. If the progression of the PD subtyping field is conditional on this flawed application of LPA, there

are concerns for the validity of the derived results. This is not to say that previous findings are necessarily invalid, but simply that there is a strong risk of bias.

If the field of PD subtyping is to progress meaningfully, these two schools of thought need to be joined. Firstly, the validity of Jankovic et al.'s (1990) TD and PIGD classifications need to be evaluated. If these classifications prove to be robust and valid, this would provide both a strong base for further development of subtyping methods, as well as an approximate 'gold standard' against which to compare new methods. Secondly, the application of LPA in PD subtyping needs to be modified so that the statistical assumptions align with the substantive understanding of the disease. Once the outcomes of this future research are known, the PD subtyping field can begin to apply consistent and valid methodologies to progress the understanding of Parkinsonian heterogeneity.

The following chapter evaluates a key concern with Jankovic et al.'s (1990) TD & PIGD subtypes: the stability of subtype membership over time. The results of this evaluation are then used to inform the development of a new approach to motor subtyping in PD.

2 Chapter 2 – Progression of Tremor and Postural Symptom Severity

2.1 Introduction

2.1.1 Stability of Subtype Membership over Time

Since their introduction, the tremor dominant (TD) and postural instability and gait difficulty dominant (PIGD) subtypes have been used extensively in Parkinson's disease (PD) research. A key concern, however, is that they have since been shown to have poor longitudinal stability (Alves, Wentzel-Larsen, Aarsland, & Larsen, 2005; Baba et al., 2012). To date, five longitudinal studies in PD have reported TD/PIGD classification counts at both the beginning and end of the study period. All five reported an increase in the proportion of people classified as PIGD and a decrease in the proportion of people classified as TD.

2.1.1.1 Alves, Larsen, Emre, Wentzel-Larsen, and Aarsland (2006)

The earliest study reporting the progression of motor subtypes over time was conducted by Alves et al. (2006). A cohort of 171 non-demented individuals with idiopathic PD, confirmed by structural imaging, were recruited and followed for eight years (Alves et al., 2006). The cohort completed a motor and cognitive evaluation at four-year intervals (baseline, three years, eight years) to determine cognitive performance, including dementia, and motor subtype status.

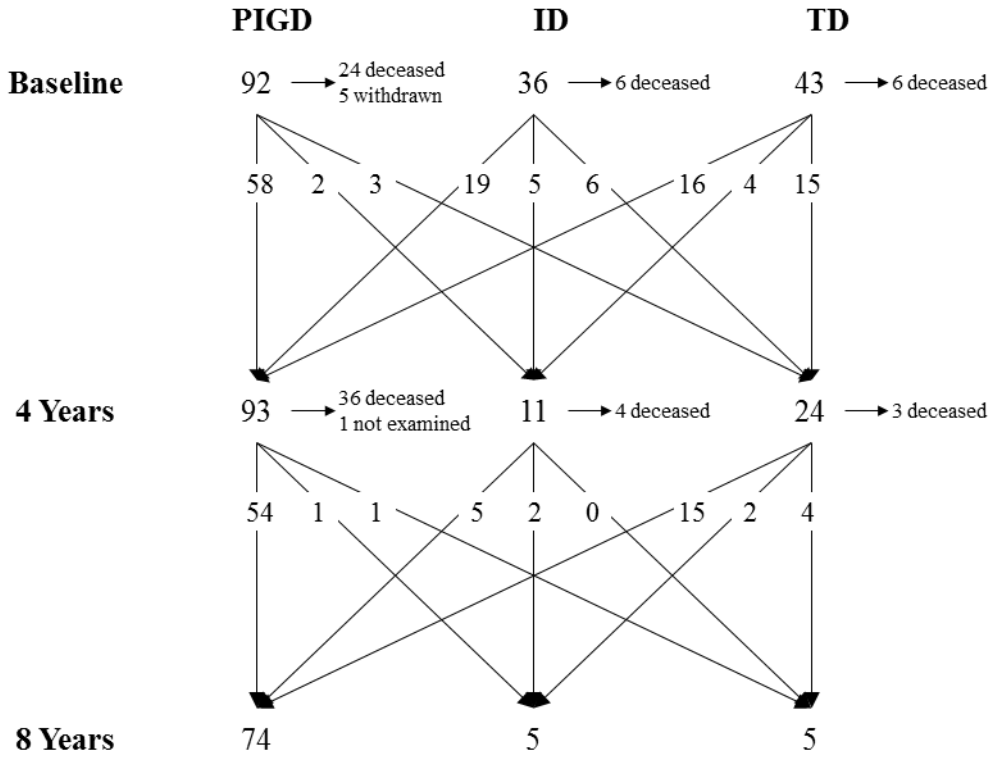
Of the 171 participants at baseline, 92 (53.8%) were classified as PIGD, 43 (25.1%) as TD, and 36 (21.1%) as indeterminate. A clear difference in mortality between the subtypes was evident when the cohort was re-assessed at four years. Of the 36 individuals who had died, 24 belonged to the PIGD subtype and 6 each to the TD and indeterminate subtypes. Each subtype group also showed clear differences in how they progressed throughout the assessment period. Alves et al. (2006) reported that from baseline to the four-year follow-up,

16 TD individuals and 19 indeterminate individuals had transitioned to PIGD subtype membership. In contrast, only 3 PIGD individuals had transitioned to the TD subtype and 2 to the indeterminate subtype. This transition to the PIGD subtype, and subsequent death, was also evident at the eight-year follow-up. Of the 43 individuals who had died in the period following the four-year assessment, 36 were from the PIGD subtype alone (with 3 belonging to TD and 4 to indeterminate). Alves et al. (2006) further identified that 15 TD individuals and 5 indeterminate individuals had transitioned to the PIGD subtype, in contrast to just 1 PIGD individual transitioning to either TD or indeterminate.

The majority of the cohort observed by Alves et al. (2006) over the eight years of follow-up eventually transitioned to the PIGD motor subtype. Further, more individuals in the PIGD subtype died than transitioned to another subtype. If this is true of the wider PD population, it may indicate that the PIGD subtype is simply a stage in the progression of PD that precedes death, rather than a distinct subtype of PD. The progression of participants between subtypes throughout the study is depicted in Figure 1. This figure shows the number of individuals belonging to each subtype at each time point, as well as the number of participants who transitioned to each of the other subtypes at each time point.

Figure 1

Progression of Participants between Motor Symptom Subtypes in Alves et al. (2006)



Note. Adapted from Alves et al. (2006)

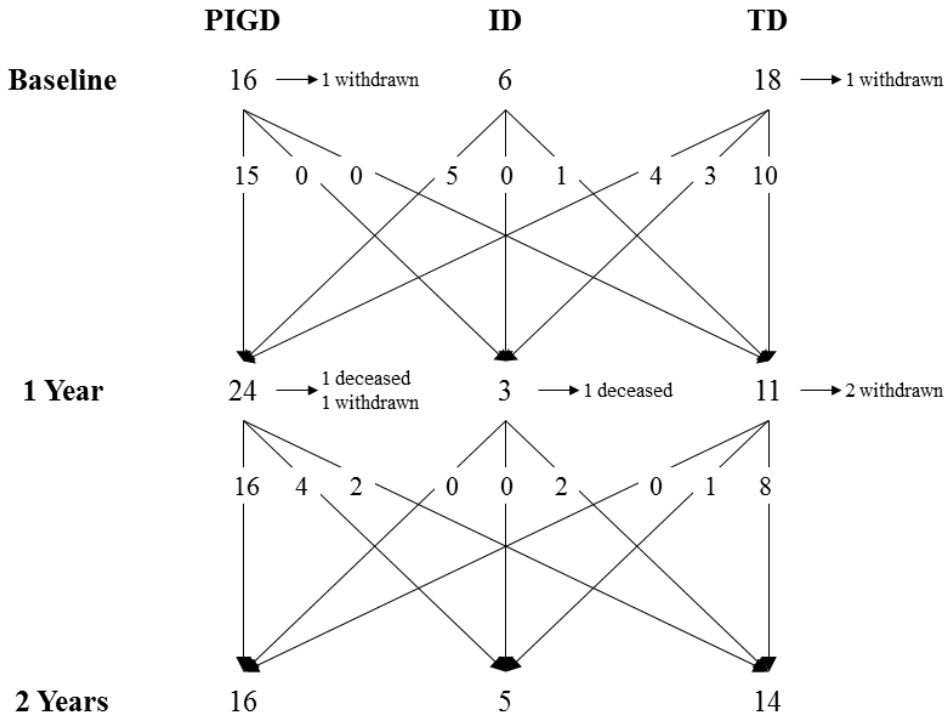
2.1.1.2 Burn et al. (2006)

Alves et al.'s (2006) observation of a common progression to the PIGD subtype has been supported by other longitudinal studies of motor subtypes in PD. The next study was conducted by Burn et al. (2006), who assessed the relationship between cognition and motor subtype over two years. A cohort of 40 non-demented individuals who met the diagnostic criteria for idiopathic PD were followed for two years. Participants completed a series of cognitive and motor assessments each year (baseline, year one, year two). Of the 40 participants at baseline; 16 (40%) were classified as PIGD, 18 (45%) as TD, and 3 (15%) as indeterminate. Of the 35 participants at the final (three year) assessment, 16 (45.71%) were classified as PIGD, 14 (40%) as TD, and 5 (14.29%) as indeterminate. Over the course of the study, the proportion of individuals classified as PIGD increased and the proportion classified

as TD decreased (see Figure 2). While these changes in proportion (approximately five percent) are relatively small compared to those reported by Alves et al. (2006), this is likely due to the much shorter length of follow-up.

Figure 2

Progression of Participants between Motor Symptom Subtypes in Burn et al. (2006)



Note. Adapted from Burn et al. (2006)

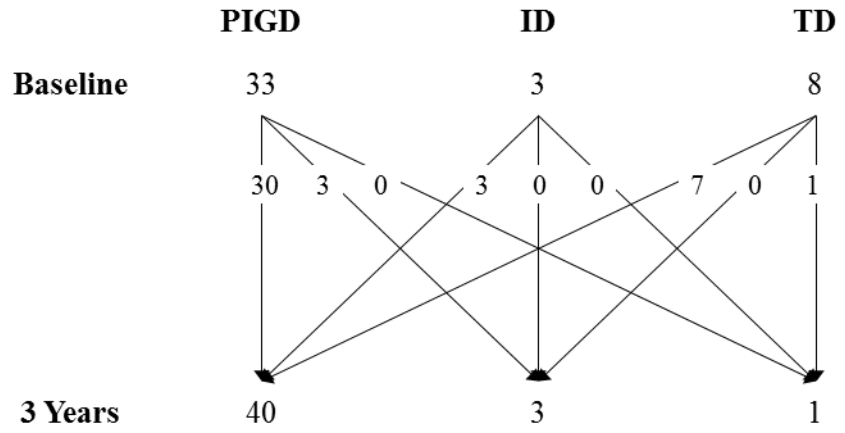
2.1.1.3 Baba et al. (2012)

Similar to the findings of Alves et al. (2006), Baba et al. (2012) reported a clear and consistent progression to the PIGD subtype over time. Baba et al. (2012) assessed the predictive relationship between olfactory performance and the subsequent development of dementia. A sample of 44 non-demented individuals with a diagnosis of PD meeting established medical criteria were recruited and followed for three years. A series of cognitive, motor, olfactory, and neuroimaging assessments were administered to participants at baseline and three-years later. Of the 44 participants at baseline; 33 (75%) were classified as

belonging to the PIGD subtype, 8 (18.18%) as TD, and 3 (6.82%) as indeterminate. At the three-year follow-up, all but one of the TD and indeterminate individuals had transitioned to the PIGD subtype. Only three PIGD individuals transitioned to a different subtype (indeterminate; see Figure 3).

Figure 3

Progression of Participants between Motor Symptom Subtypes in Baba et al. (2012)



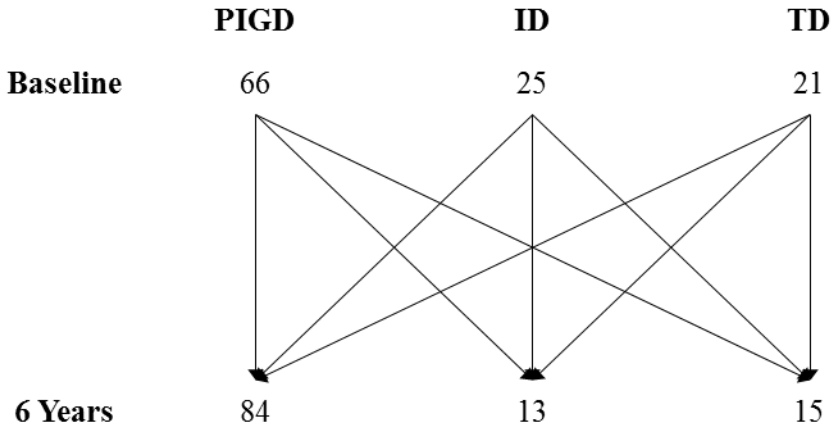
Note. Adapted from Baba et al. (2012)

2.1.1.4 Nabli et al. (2015)

Nabli et al. (2015) also reported a consistent progression to the PIGD subtype. Nabli et al. (2015) assessed whether the motor and functional decline of individuals with a LRRK2 G2019S mutation and idiopathic PD differed from those without the mutation. Participants were assessed at study entry (baseline) and again six years later. No significant difference in the progression of motor symptoms or motor subtypes over six years between the two genetic groups was found (Nabli et al., 2015). Therefore, the subtype counts for the combined groups were reported. At baseline; 58.93% of the sample belonged to the PIGD subtype, 22.32% to the TD subtype, and 18.75% to the indeterminate subtype. At the six-year follow-up, 75% were classified as PIGD, 13.39% as TD, and 11.61% as indeterminate (see Figure 4).

Figure 4

Progression of Participants between Motor Symptom Subtypes in Nabli et al. (2015)



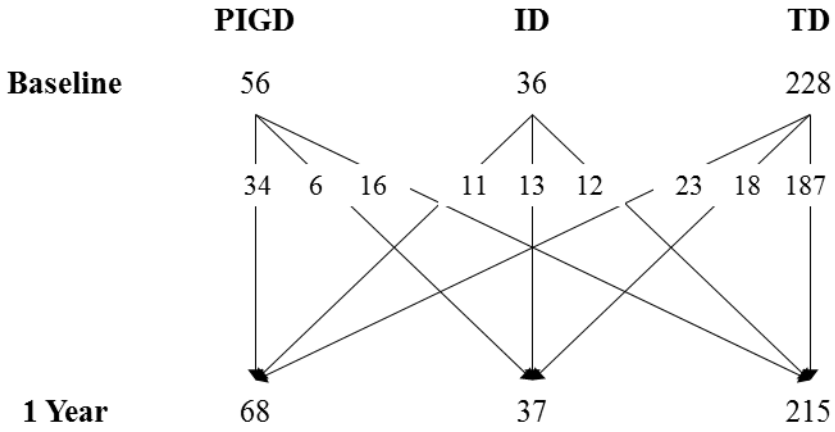
Note. Adapted from Nabli et al. (2015). Transition counts within each subtype were not reported

2.1.1.5 Simuni et al. (2016)

Simuni et al. (2016) assessed the stability of motor subtype over a one-year period in de novo PD individuals. While the proportions of subtype changes were much less marked than the other studies, Sinuni et al. (2016) reported an increase in PIGD prevalence and a decrease in TD prevalence. At baseline; 17.5% of individuals were classified as PIGD, 71.25% as TD, and 11.25% as indeterminate. At the one-year follow-up assessment; 21.25% were classified as PIGD, 67.19% as TD, and 11.56% as indeterminate (see Figure 5). Further, the markedly different number of individuals in the TD and PIGD subtypes at baseline are consistent with a progression from TD to PIGD membership. All individuals in this study were newly diagnosed (de novo), and so the majority may not have sufficiently progressed in the disease to have transitioned from the TD to the PIGD subtype.

Figure 5

Progression of Participants between Motor Symptom Subtypes in Sinuni et al. (2016)



Note. *Adapted from Sinuni et al. (2016).*

Taken together, the findings of these five longitudinal studies indicate that PD motor subtype, as it is currently conceptualised, is not stable over time. Furthermore, there appears to be a common progression to the PIGD subtype over the course of the disease. These findings indicate that the current conceptualisation and application of motor subtypes in PD is of concern. If motor symptom subtypes are not reliable or predictable, they have little clinical and prognostic value for individuals, clinicians, or researchers. This is critical, as several studies have argued that the PIGD subtype is a risk factor for cognitive decline in non-demented individuals with PD (Alves et al., 2006; Burn et al., 2006). If motor subtype is not stable over time, then individuals presenting as TD would still be at risk, given the high likelihood of their eventual transition to the PIGD subtype.

2.1.2 Stability of Motor Symptom Severity over Time

A potential source of longitudinal instability in motor subtypes may be the ways in which the motor symptoms themselves (tremor and postural) change over time. If both types of symptoms progress similarly, and symptom differences are stable over time, then it follows that the motor subtypes themselves would remain stable. If, however, the motor

symptoms progress in severity at different rates over time, then the differences in severity between symptoms (and thus the resulting motor subtype) would not be stable.

2.1.2.1 Louis et al. (1999)

To date, only two studies have reported changes in tremor and postural symptoms over an extended period. Louis et al. (1999) assessed the changes in motor symptom severity, cognition, and experiences of daily living over eight years in a community-based cohort of 237 individuals with PD. Participants completed a series of motor, cognitive, and daily living assessments annually. Using generalised estimating equations, the authors found that postural motor symptoms changed significantly on a yearly basis, but that tremor symptoms did not. However, the use of generalised estimating equations in this context provides a rather limited perspective of the progressive nature of PD and the findings should be treated with caution.

The purpose of generalised estimating equations is to estimate an average rate of change in the symptom of interest, while correcting for the correlated nature of repeated measurements (Hanley, 2003). This approach attempts to distil the symptom change within a sample into a single statistic, which may not accurately represent the actual change that occurred. For example, the use of an average rate of change does not indicate whether the actual rate of change was different throughout the course of the study. That is, were the changes in severity relatively small at the start of the study but much larger at the end? By reducing all change into a single representative number, an average rate of change does not give a meaningful picture of how symptoms progressed over time.

Another issue to be considered is that the authors only tested whether the rates of symptom change were significantly different from zero, not whether one was significantly smaller or larger than the other (Louis et al., 1999). The authors never explicitly tested whether the rate of change in postural symptoms was significantly different from the rate of

change in tremor symptoms. Given the possibility of Type I/II errors, the authors may not be able to say with statistical confidence whether the progression of tremor symptoms was significantly different from the progression of postural symptoms. Overall, the results of Louis et al. (1999) provide initial evidence of possible differences in the progression of tremor and postural symptoms over time in PD. However, their use of generalised estimating equations limits the strength and applicability of this evidence.

2.1.2.2 Vu, Nutt, and Holford (2012)

The second study reporting changes in tremor and postural symptoms over time was conducted by Vu et al. (2012). The authors re-analysed the data from the cohort used by Jankovic et al. (1990) for the original proposal of the TD and PIGD subtypes (Vu et al., 2012). Seven hundred and ninety-five individuals with diagnosed PD participated in a randomised controlled trial. Initially untreated at study entry, participants were randomised to 1 of 4 intervention groups: placebo, selegiline, tocopherol, or both. After two years, selegiline was identified as the most effective of the interventions at delaying the need for antiparkinsonian medication (Vu et al., 2012). All participants in the study were then moved to the selegiline intervention group.

Vu and colleagues (2012) used a self-developed pharmacodynamic approach to estimate and apply models of symptom change over time. The authors used symptom "half-life" as indicative of the rate of change in symptom severity over time. That is, the length of time taken for the symptom to become half as severe as the possible maximum score (similar to the definition for chemical decay). The authors identified a slightly slower rate of change in tremor symptoms (half-life of 3.9 years) compared to postural symptoms (3.1 years). However, as with Louis et al. (1999), the authors did not statistically test for a difference between the two. It is, therefore, possible that the observed differences may be within the range of measurement or statistical error.

2.1.3 Statistical Modelling of Change over Time

While novel, the previous explorations by Louis et al. (1999) and Vu et al. (2012) were both limited in the depth with which they could characterise the changes in tremor and postural symptoms in PD. Neither study was able to present a clear model describing the progression of symptom severity over time. Further, these studies were not able to present evidence that tremor and postural symptoms had significantly different slopes (or rates) of change over time. Both need to be established accurately to assess whether a consistent progression to the PIGD subtype is due to differential rates of progression in symptom severity.

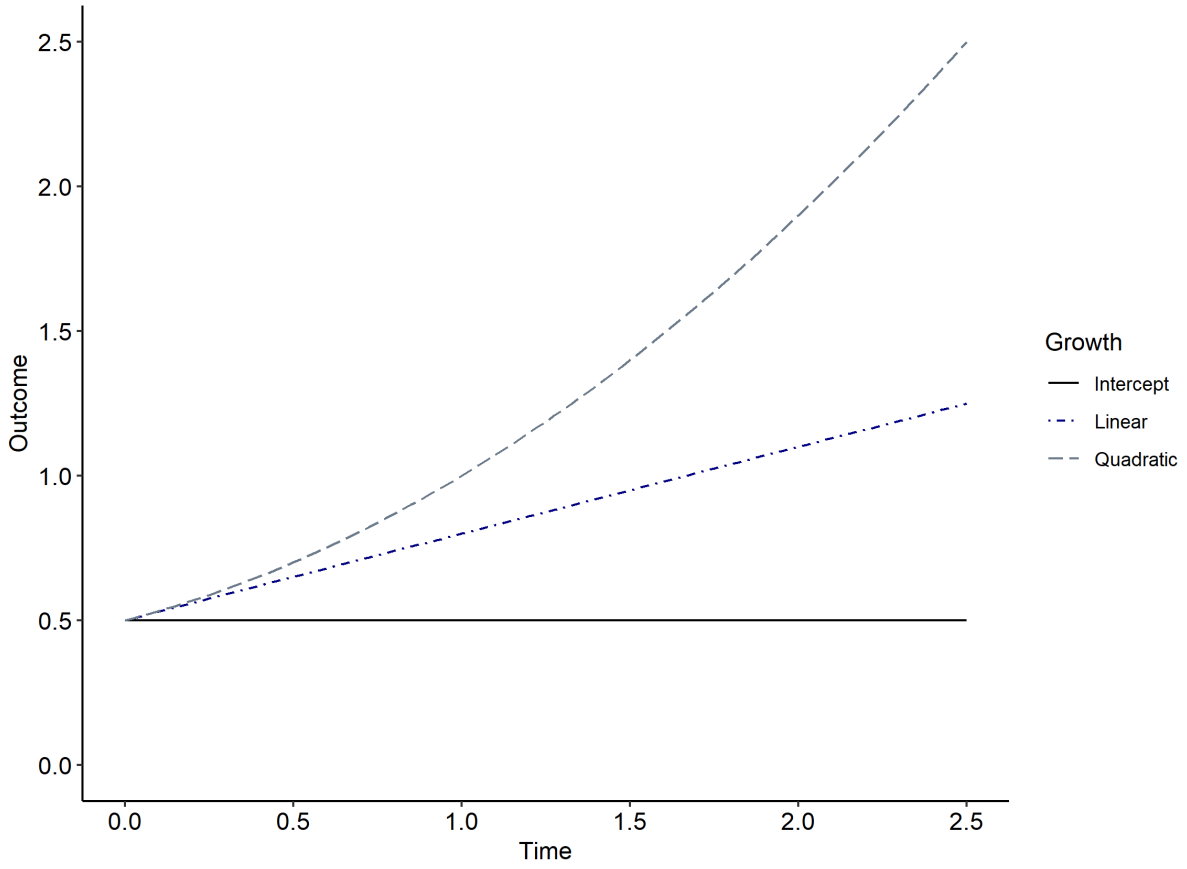
A more appropriate form of longitudinal analysis is latent growth modelling. Latent growth modelling uses latent variables as a means of describing and quantifying how individuals change over time. As already discussed in the context of the findings of Louis et al. (1999), a single statistic is not sufficient to describe a concept as complex as growth (change over time). To address this, latent growth modelling deconstructs growth over time into two components: intercept and slope.

The intercept component of latent growth modelling represents the severity of symptoms at first (baseline) measurement. This is the 'starting point' for the individual, with the subsequent components describing how the individual changes from this first measurement. The slope component represents the linear change in severity from the baseline measurement (intercept) at each successive follow-up. The slope component assumes that symptom severity changes by the same amount at each follow-up. This may not be realistic in PD, where the rate of neurological degeneration (and associated changes in symptom severity) is not constant over time (Pirker et al., 2002). If this is the case, a quadratic component can be added, which allows for the amount of symptom change at each follow-up

to be different. The model of growth implied by each of these components is depicted in Figure 6.

Figure 6

Change over Time as Conceptualised by Latent Growth Modelling



By applying these models of growth to the observed measurements and comparing the fit of each model, the type and rate of change for a given symptom can be more accurately determined. Latent growth modelling is also uniquely suitable for hypothesis testing with multiple symptoms. Once the appropriate model of change for each symptom has been determined, the growth of both symptoms can be simultaneously estimated in the same model. This allows for the statistical assessment of whether the changes in the two symptoms are related (i.e., whether the changes of one symptom influence the changes of another), and whether the rates of change are significantly different between the two.

To date, only one study has used latent growth modelling to explore the progression of PD motor symptoms. Zahodne et al. (2012) assessed the progression of overall motor severity (the mean of all symptom severity scores), apathy, and depression in 186 individuals with idiopathic PD over 18 months (baseline, 6 months, 18 months). Results revealed that both overall motor severity and apathy increased at a linear (constant) rate, while depression increased in severity at a quadratic (increasing) rate. One key, methodological limitation of this study was the assessment of *overall* motor severity. Given the considerable heterogeneity in PD motor symptoms, it is possible that participants could present with similar overall motor severity scores but have markedly different individual symptom scores. This was observed by Lord et al. (2014), in their study of the relationship between cognition and gait in PD in 121 individuals. Participants were subtyped as TD or PIGD, and the mean of all motor severity scores for these two groups was reported (Lord et al., 2014). The subtyping into TD and PIGD groups would suggest that the severity of these two symptoms were markedly different between the two groups. Despite this, there was no significant difference in overall motor severity, $t(106) = -1.16$ [-6.23, 1.63], $p = .249$ (Lord et al., 2014). In light of this, collapsing motor symptoms into a single 'representative' score may obscure important individual differences in the patterns of change over time.

The present study used latent growth modelling to assess the longitudinal progression of tremor and postural symptoms in people with PD without dementia over a six-year period, to determine whether both symptoms progress at the same rate (Aim 1) and whether the changes in the two symptoms are related (Aim 2).

2.2 Methods

2.2.1 Participants

Data were collected by the ongoing ParkC study (Loftus et al., 2015), which has approval from the Human Research Ethics Committee of Curtin University (HR158/2013). A self-referred, community-based cohort was recruited. Assessments took place at a university clinic or within the participant's home, at their discretion. Written informed consent was obtained from each participant at each assessment (including follow-up). A series of cognitive, mood, and motor assessments was conducted every two years during a participant's medication 'on' state. The full testing protocol has previously been described (Loftus et al., 2015). Inclusion in the study required a diagnosis of PD by a neurologist or geriatrician in accordance with the United Kingdom Parkinson's disease Brain Bank Criteria (Hughes, Daniel, Kilford, & Lees, 1992). As recruitment is ongoing, not all participants had completed all follow-up assessments. The current sample comprised 236 participants: with data from 236 participants at Time 1, 154 at Time 2, 95 at Time 3, and 36 at Time 4. While markedly fewer individuals had been assessed at Time 4, latent growth modelling is robust in accounting for incomplete follow-up data (McArdle & Hamagami, 1992). Sample demographics are reported in Table 1.

2.2.2 Measures

Motor symptoms were assessed using the Movement Disorders Society revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS; Goetz et al., 2008). The MDS-UPDRS is a 50-item clinical assessment of PD severity. The assessment comprises four sections, each evaluating a different dimension of PD symptomology. Part III of the MDS-UPDRS, used in the present study, is a physical assessment of motor symptom severity conducted by a trained clinician or researcher. The assessment requires the participant to complete 33 different movements; assessing the severity of resting and kinetic tremor,

postural instability, rigidity, and akinesia/bradykinesia. The tremor and postural symptom scores used in the present study were derived using the criteria proposed by Jankovic et al. (1990), and revised by Stebbins et al. (2013). This method requires calculating the mean of the MDS-UPDRS items assessing tremor and postural severity, respectively. These mean scores then fall in the same range as the MDS-UPDRS items. The mean tremor and postural scores can range from 0 to 4, with 0 indicating no severity and 4 indicating maximum severity (Stebbins et al., 2013). The mean postural and tremor symptom severity scores from these subscales, and the resulting TD/PIGD subtypes at each time point are presented in Table 1. Also presented is the Levodopa Equivalent Dose (LED; Tomlinson et al., 2010).

Table 1
Sample Demographics at Four Time Points Over 6 Years (N = 236)

	Time 1 (*N* = 236)	Time 2 (*N* = 154)	Time 3 (*N* = 95)	Time 4 (*N* = 36)
Sex				
Male	157	108	62	17
Female	79	46	33	19
Age	66.05 (9.57)	66.36 (9.77)	66.64 (9.35)	65.83 (8.86)
Disease Duration (Months)	65.47 (57.96)	89.59 (57.89)	108.81 (53.29)	125.72 (50.98)
Mean LED	607.31 (485.45)	743.62 (502.71)	912.67 (557.65)	989.52 (577.81)
Mean Tremor	0.66 (0.47)	0.66 (0.52)	0.73 (0.48)	0.67 (0.49)
Mean Postural	0.66 (0.60)	0.79 (0.68)	1.04 (0.85)	1.09 (0.82)
TD Subtype	122 (51.69%)	70 (45.45%)	35 (36.84%)	14 (38.89%)
PIGD Subtype	87 (36.86%)	72 (46.75%)	50 (52.63%)	19 (52.78%)
Indeterminate	22 (9.32%)	10 (6.49%)	8 (8.42%)	3 (8.33%)

Note. LED = Levodopa Equivalent Dose (Tomlinson et al., 2010); TD = Tremor Dominant; PIGD = Postural Instability and Gait Difficulty

2.2.3 Data Analysis

Data were analysed using latent growth curve modelling in Mplus version 8.2. The mean tremor and postural scores were not normally distributed (positively skewed), so robust maximum likelihood estimation was used. Due to the nature of a longitudinal study with ongoing recruitment, not all participants had completed all follow-up assessments. Full information maximum-likelihood was used to account for some participants' lack of follow-up data, which has been shown to be reliable in latent growth modelling with large amounts of attrition (McArdle & Hamagami, 1992).

However, it should be noted that this method assumes that there are no unobserved, systematic, reasons for missingness; it assumes that the responses are missing at random

(MAR; Enders, 2011). To test whether this assumption has been met, the model should be re-fit assuming that the data are not missing at random (MNAR). If the results do not differ between the MAR and MNAR models, then the assumption has been met. Two MNAR models will be tested, each assuming a different mechanism for the missing responses: a selection model, and a pattern-mixture model. The selection model implies that an individual's missing responses are dependent on their previously observed responses (Enders, 2011). In contrast, the pattern-mixture model implies that the model parameters differ for each pattern of missing responses (Enders, 2011). For the present analysis, the best-fitting growth models were re-fit under the MNAR assumption and their estimated means compared, if these estimated means do markedly differ, then the assumption has been met.

Model fit was assessed using the Tucker and Lewis Index, threshold: > .99 (Hu & Bentler, 1999); the Comparative Fit Index, threshold: > .99 (Bentler, 1990); and the Root Mean Square Error of Approximation, threshold: < .05 (Steiger, 2007). To compare the fit of models, chi-square difference tests were used. These difference tests were supplemented with visual inspection of the Bayesian Information Criterion (lower values indicate a comparatively better-fitting model).

The linear slope of the latent growth modelling was parameterised using fixed time scores of 0, 2, 4, and 6 for Times 1 to 4. The linear time scores are used to represent the amount of time between each assessment, and the unit of time being used (i.e., two, four, and six years). The quadratic slope was parameterised using the squared time scores (0, 4, 16, 36). The quadratic time scores are used to model the acceleration in change between each assessment (i.e., the amount of change that occurred beyond the linear change).

Tremor symptoms and postural symptoms were first modelled independently, to determine the best fitting model to describe their growth processes (intercept-only, linear, or

quadratic). Next, levodopa equivalent dose at each time point was entered to assess the impact of antiparkinsonian medication on symptom progression. Following this, using the best fitting growth models, both symptoms were modelled simultaneously to assess whether the baseline or growth in severity of one symptom was related to the baseline or growth in severity of the other. Finally, the effects of sex, baseline age, and baseline disease duration on symptom progression were assessed.

2.3 Results

2.3.1 Sample Characteristics

When visually assessing the sample demographics at each assessment time, there were clear differences in age and sex. At the baseline assessment, there were almost twice as many males in the sample as females. Yet, at the final assessment six years later the proportions were almost equal, suggesting differential rates of mortality and loss to follow-up. It should be noted that the large majority of 'loss to follow-up' is due to individuals being recruited at different times and not yet due for follow-up. Of further interest are the different amounts of increase in the mean disease duration and age of individuals in the sample. Disease duration increased as expected between assessments (i.e. approximately two years difference). In contrast, the sample average age remained stable at (approximately) 66 years. This could imply the presence of a (somewhat) common point of mortality or high risk, as the average age of the sample was stable regardless of the time of disease diagnosis. The relationships between these items and the changes in motor symptoms further explore these points.

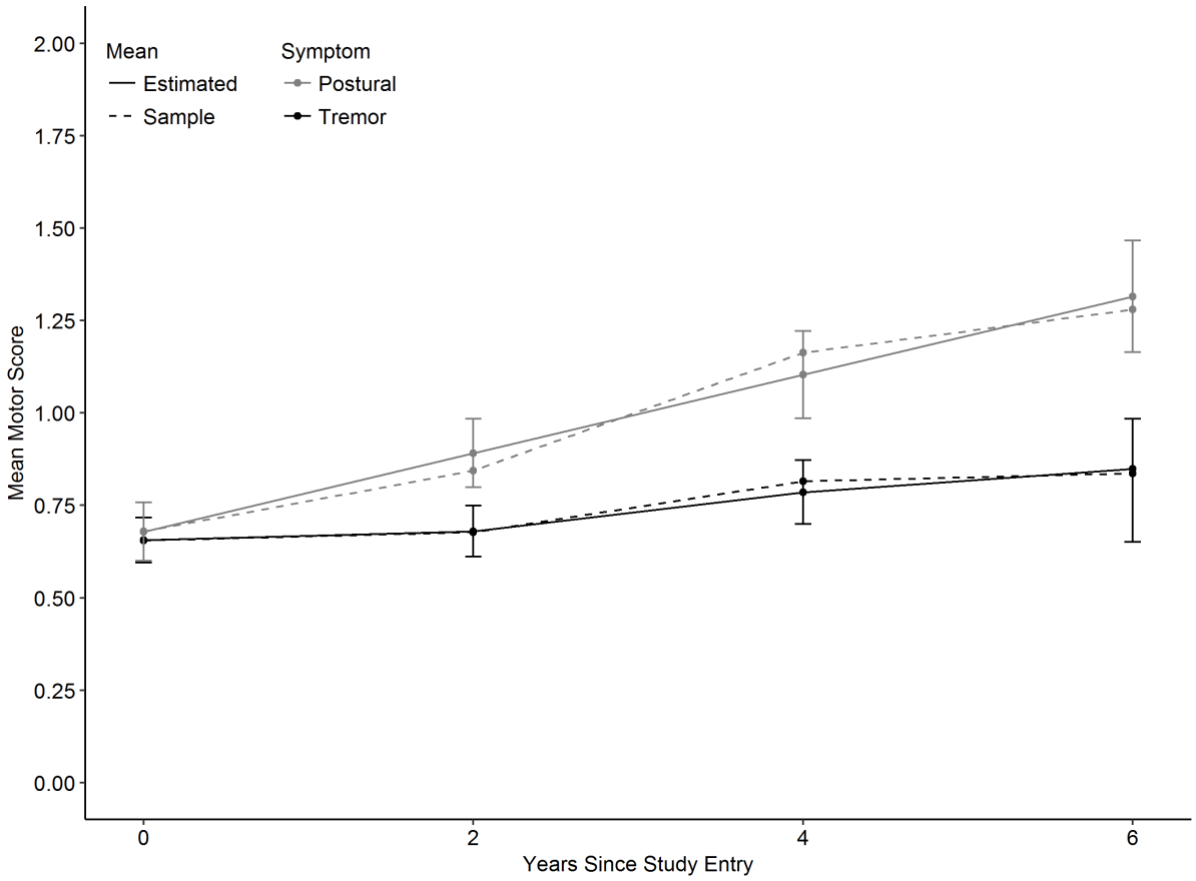
2.3.2 Unconditional Growth Models

After accounting for individual variation due to antiparkinsonian medication, tremor symptoms revealed the best fit with a linear model of change. Before controlling for antiparkinsonian medication, the model estimated that individuals entered the study with an

average tremor score of 0.65, which varied significantly between individuals. The model indicated that this average tremor score increased by a statistically significant 0.03 points per year. After accounting for antiparkinsonian medication at each time of measurement, the model estimated that the actual change in tremor severity was a non-significant 0.004 points per year and that this did not significantly vary between individuals.

Postural symptoms also revealed the best fit to the data with a linear model of change but were not significantly affected by antiparkinsonian medication at each time of measurement. The model estimated that individuals presented with a baseline mean postural symptom score of 0.68, which increased by a non-significant 0.11 at each successive visit. In contrast to the tremor model of change, when antiparkinsonian medication at each time point was entered the model fit deteriorated. Further, there were no significant relationships between antiparkinsonian medication and postural symptom severity at any time point. This indicates that the longitudinal progression of postural symptoms was not related to changes in antiparkinsonian medication. Graphs of the sample and estimated means for tremor and postural symptoms, with confidence intervals, are presented in Figure 7.

Figure 7
Sample and Estimated Mean Motor Symptoms with 95% Confidence Intervals



The two growth processes (tremor and postural as linear) were then jointly estimated to determine whether tremor and postural symptoms were related in terms of initial severity and/or rates of change (i.e., progression of symptoms). Fit deteriorated significantly (Comparative Fit Index and Tucker and Lewis Index below thresholds) and there were no correlations between growth factors for either symptom. This suggests that the severity and rates of change of tremor and postural symptoms were unrelated. Covariates were therefore explored for each model separately. Fit statistics for all tested models are reported in Table 2

Table 2
Fit Statistics for Tremor and Postural Growth Models

Model	BIC	χ^2 (*df*), *p*	$\Delta\chi^2$ (*df*), *p*	RMSEA	CFI	TLI
Tremor						
Tremor Intercept	571.91	17.96 (8), *p* = .022		.07 [.03, .12]	.92	.94
Tremor Linear	568.55	5.72 (5), *p* = .335	11.58 (3), *p* = .009	.02 [.00, .10]	.99	.99
Tremor Quadratic	580.50	9.37 (2), *p* = .009	2.22 (3), *p* = .567	.12 [.05, .21]	.94	.83
Tremor Linear (LED)*	1287.59	18.42 (17), *p* = .363	12.58 (12), *p* = .400	.02 [.00, .06]	.99	.99
Postural						
Postural Intercept	895.94	125.23 (8), *p* < .001		.25 [.21, .29]	.63	.73
Postural Linear*	816.09	8.15 (6), *p* = .228	312.54 (2), *p* < .001	.04 [.00, .10]	.99	.99
Postural Quadratic	821.10	7.86 (5), *p* = .164	.43 (1), *p* = .512	.05 [.00, .11]	.99	.99
Postural Linear (LED)	1546.27	28.41 (18), *p* = .056	19.92 (12), *p* = .069	.05 [.00, .08]	.95	.94
Joint						
Tremor Linear (LED) & Postural Linear	2123.431	79.92 (51), *p* = .006		.05 [.03, .07]	.93	.92

Note. * Selected as final model. LED = Levodopa Equivalent Dose, BIC = Bayesian Information Criteria, RMSEA = Root Mean Squared Error of Approximation, CFI = Confirmatory Fit Index, TLI = Tucker and Lewis Index

2.3.3 Conditional Growth Models

Separate, conditional growth models were estimated to account for the significant variance seen in baseline tremor and postural severity, and in tremor rate of change. Sex (male=1, female=0), age at baseline, and disease duration at baseline were entered as predictors of the tremor intercept and slope, and as predictors of the postural intercept and slope.

2.3.3.1 Sex

Males presented with more severe initial tremor symptoms than females (standardised B = .40 [.13, .68], p = .004). However, there was no significant difference between sexes in initial postural severity or in the rate of change of either symptom.

2.3.3.2 Age

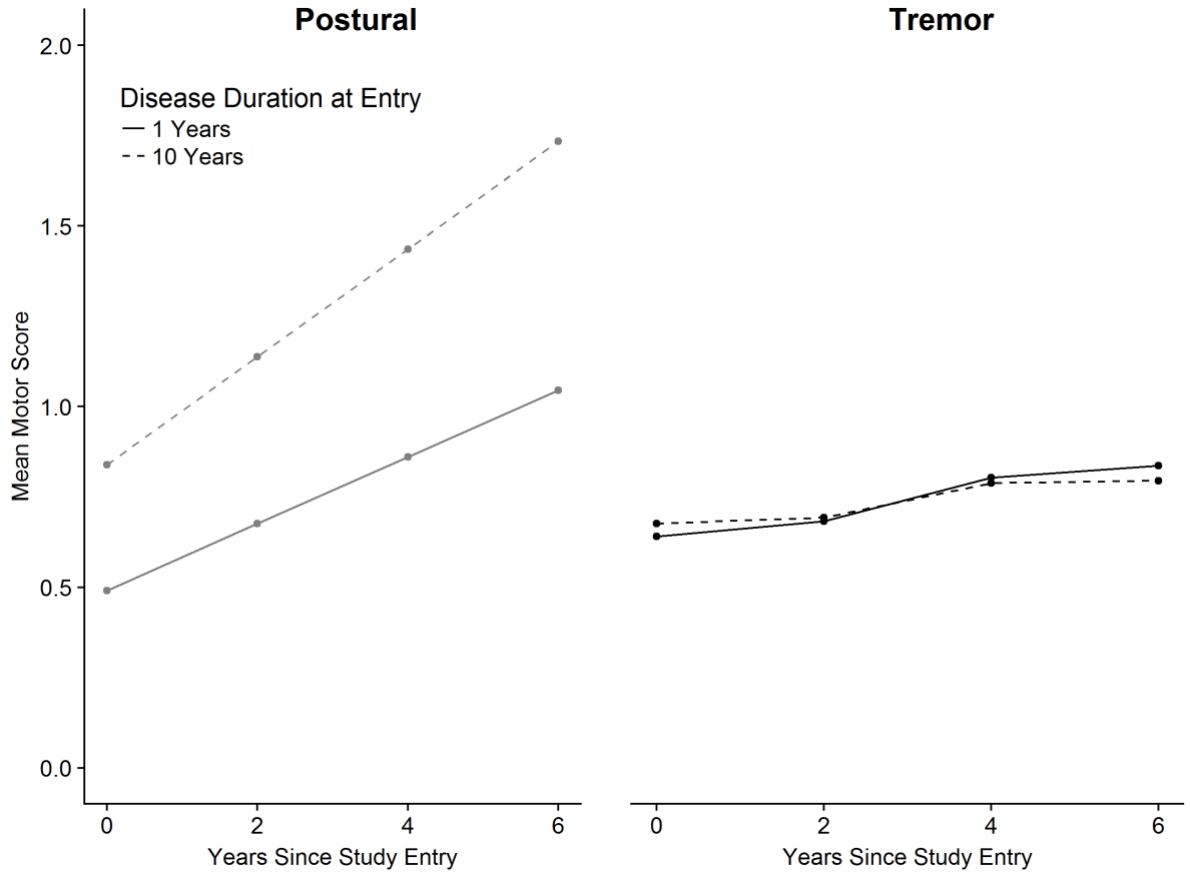
Older individuals presented with more severe tremor (standardised B = .20 [.07, .33], p = .002) and more severe postural symptoms (standardised B = .19 [.17, .21], p = .002) at baseline, and showed a faster rate of change in postural symptom severity (standardised B = .68 [.32, .99], p < .001).

2.3.3.3 Disease Duration

Individuals with longer disease duration at baseline presented more severe postural symptoms (standardised B = .32 [.17, .46], p < .001) compared to those with shorter duration at entry. These individuals also demonstrated a faster progression of postural symptom severity (standardised B = .44 [.07, .82], p = .020). There was little difference in tremor severity between those with longer disease duration and those with shorter disease duration. These findings are depicted in Figure 8, where the estimated means for individuals who presented with a duration of 1 year are compared to those of individuals who presented with a duration of 10 years.

Figure 8

Model Estimated Disease Duration Differences in Motor Symptom Severity and Progression



2.3.4 MAR Assumption Testing

Prior to discussing the implications of these findings, it is important to first assess whether the assumption of missing at random (MAR) data is tenable. If the missing data are not missing at random (as assumed by the full-information maximum likelihood approach), then the presented results may not be reliable. The final postural and tremor growth models were re-fit under two missing not at random (MNAR) assumptions: the selection model and the pattern-mixture model. The model-estimated means under the MNAR assumptions were highly consistent with the estimated means under the MAR approach for both the tremor growth model (see Figure 9) and the postural growth model (see Figure 10). This implies that the assumption of MAR data was appropriate.

Figure 9
Model Estimated Tremor Mean Under MAR and MNAR Assumptions

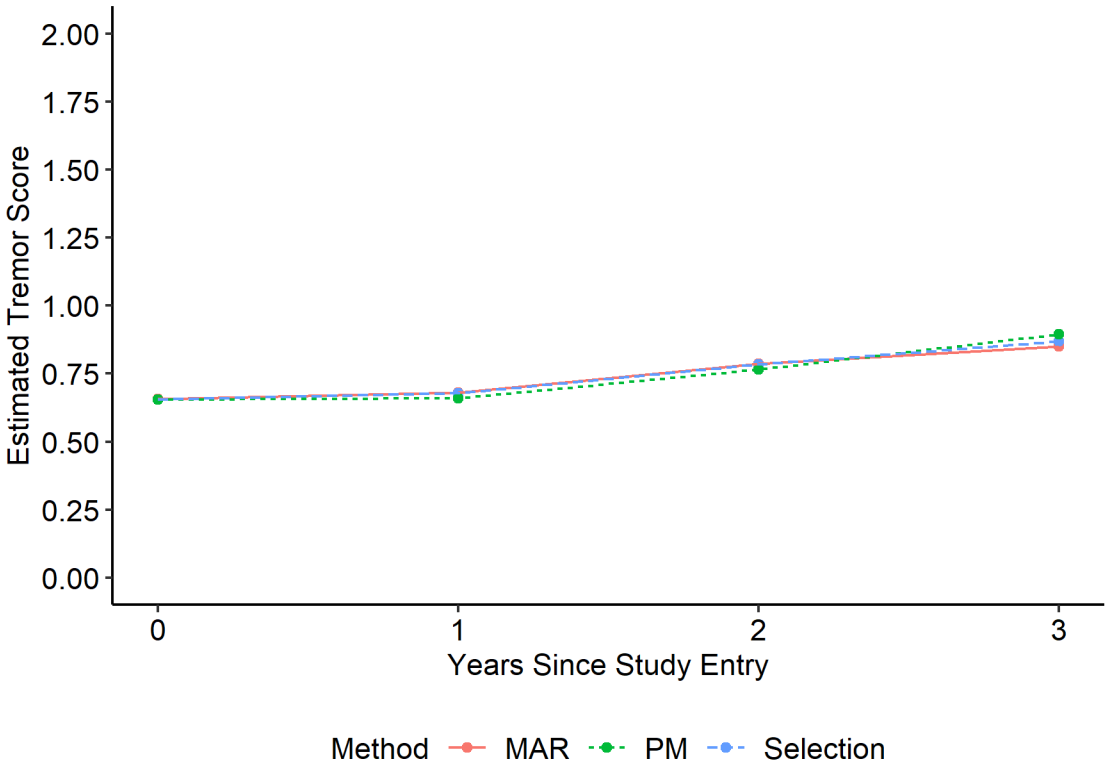
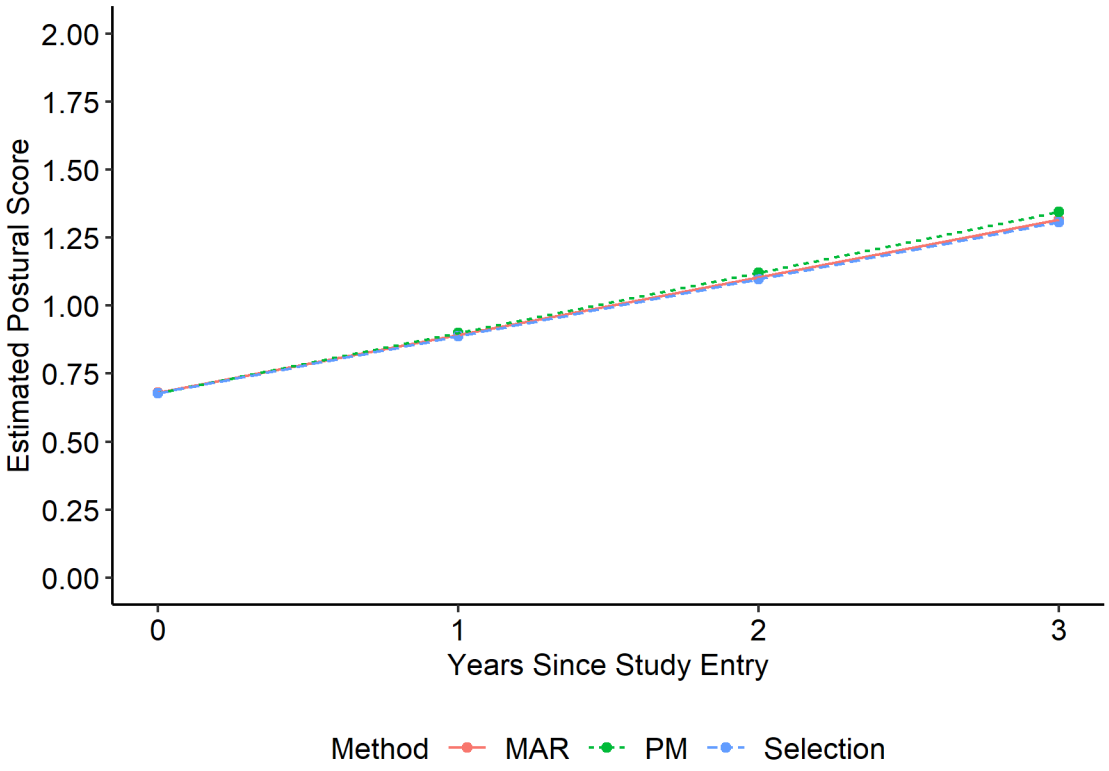


Figure 10
Model Estimated Postural Mean Under MAR and MNAR Assumptions



2.4 Discussion

This study sought to determine whether tremor and postural symptoms progress at the same rate (Aim 1) and whether the changes in these symptoms are related (Aim 2). It also assessed whether these changes were a function of disease duration, age, or sex. Results showed that, after controlling for antiparkinsonian medication, tremor and postural symptoms did not progress at the same rate. Tremor symptoms did not significantly change over time, although the rate of progression varied significantly between individuals. In contrast, postural symptom severity deteriorated significantly over time. Given the marked differences in the rate of change of tremor and postural symptoms over six years, it should not be remarkable that the progressions in severity of these symptoms was unrelated. The progression of tremor symptoms was not related to disease duration, age at entry, or sex; although the progression of postural symptoms was significantly increased by age and disease duration. Baseline tremor symptoms were more severe in males than in females, and more severe in individuals who were older at study entry. Baseline postural symptoms were more severe in those individuals who were older and/or had longer disease durations at study entry.

The analyses showed that, over the six years of follow-up, postural symptoms became more severe, but tremor symptoms did not. If postural but not tremor symptom severity increases over time, then the dominance of tremor or postural symptoms could be driven by the duration of PD. Such a view is consistent with recent work by Pavese, Rivero-Bosch, Lewis, Whone, and Brooks (2011), who assessed changes in dopaminergic neuropathology in 10 individuals with PD over three years. Pavese et al. (2011) identified that degeneration of nigrostriatal dopaminergic structures was independent of extranigral degeneration. Furthermore, extranigral degeneration did not occur until later in the progression of the disease (Pavese et al., 2011). Tremor severity is suggested to be driven by nigrostriatal dopaminergic degeneration, and postural severity by extranigral cholinergic degeneration

(Bohnen & Albin, 2011). If, as Pavese et al. (2011) suggest, nigrostriatal and extranigral degeneration are independent, this would account for the independent progression of tremor and postural symptoms. Further, if nigrostriatal degeneration precedes extranigral degeneration, this would account for tremor and postural symptoms differing in when they become severe, with the latter presenting only in later PD.

Given that the progression of tremor and postural symptoms in our sample was unrelated, it is also plausible that other PD symptoms may have independent rates of progression. This could have implications for the accuracy of the identification of other motor subtypes, such as the akinetic-rigid subtype (Eggers, Kahraman, Fink, Schmidt, & Timmermann, 2011), and for the subtypes of mild cognitive impairment in PD (Litvan et al., 2011). If the motor and cognitive symptoms that comprise these subtypes do not share the same rates of progression, the magnitude of their difference could also be a function of disease duration, rather than of intra-individual differences in presentation. The presentation of symptom differences may be both time and disease dependent, with subtype differences only present after accounting for time differences.

While these results do not provide support for cross-sectional TD and PIGD subtypes, they do appear to support a 'staging' theory of these subtypes. Consistent with the literature presented earlier, these results suggest that the TD and PIGD subtypes are representative of different stages in the progression of PD. In other words, rather than classifying individual differences in disease presentation these subtypes appear to be classifying individual differences in disease progression. This discussion has previously been raised by Nutt (2016), who argued that, whilst there is strong evidence that PD subtypes are likely present, the literature to date has not been able to convincingly demonstrate that the TD/PIGD subtypes are biologically distinct and not stages of PD. The likelihood of a staging process should be evaluated as part of any future subtyping research.

Some methodological limitations of this study also need to be considered. All participants were assessed in the 'ON' state of their medication, which may have influenced the severity of their symptom presentation. How this affected their progression of symptom severity remains unknown. This is a common difficulty for longitudinal studies, including those of Louis et al. (1999), Vu et al. (2012), and Burn et al. (2006), among others. It should, however, be noted that dosage of antiparkinsonian medication in the current study was not related to the progression of tremor or postural symptoms, only to early (but not late) tremor severity. This would suggest that being assessed in the ON state does not explain the different trajectory of tremor and postural symptoms, despite the greater effectiveness of antiparkinsonian medication for treating tremor (Bohnen & Albin, 2011).

Beyond this, a clear concern of the present research is the declining sample size throughout the study, especially so in the final time point. The full-information maximum-likelihood (FIML) procedure used to account for this is robust to decreasing samples in growth models (McArdle & Hamagami, 1992). However, there is still the likelihood that the individuals assessed at this final time point are not representative of the remainder of the sample, and that this could bias the results in some manner. It should be noted that both the observed and estimated trajectories of motor severity (Figure 7) also demonstrated a marked difference between tremor and postural severity at the second and third assessment times, consistent with the final assessment time. This indicates that the progression of severity (both estimated and observed) at the final assessment was consistent with the measured severity at previous assessments where greater numbers of individuals had been measured. This consistency indicates that the smaller sample size at the final assessment time likely did not bias the estimation to the extent that the results were not representative of the full sample.

The results of the present study indicate that the most commonly used approach to PD subtyping currently (Jankovic et al.'s (1990) TD/PIGD groupings) has a conceptual flaw.

That is, the TD/PIGD approach to subtyping presupposes that the magnitude of difference in tremor and postural symptom severity is indicative of a different subtype of PD. However, the results of the present study indicate that this difference in severity is more likely due to the progression of the disease itself. This highlights two types of limitations in current approaches to subtyping: conceptual and methodological. Conceptual limitations could be understood as an inaccuracy in how different aspects of symptom presentation are understood to relate to different subtypes of the disease (e.g., differing levels of tremor and postural severity as the result of different subtypes). Methodological limitations could be understood as the limitations in the methods used to derive these conceptualisations or to apply them to subtyping.

While the TD/PIGD approach to subtyping has a clear conceptual flaw, this is more likely due to the methodological limitations in how these subtypes were developed. Firstly, Jankovic et al. (1990) selected the UPDRS items that they felt best assessed tremor and postural symptoms. The authors did not provide any validation that these UPDRS items were truly measuring the intended PD motor symptoms. Secondly, the authors provided no validation for their approach to subtyping. Despite these key issues, Jankovic et al. (1990) is one of the most cited studies in the PD subtyping field and these types of limitations continue to be common.

2.5 Chapter 2 – Summary

This chapter characterised the progression of tremor and postural symptom severity across a six-year period in a sample of 236 individuals. The results indicated that postural symptoms become increasingly severe over time, whereas tremor symptoms show little change over time. This finding implies that Jankovic et al.'s (1990) TD and PIGD classifications are a function of disease duration, rather than disease subtype. These findings further imply that the TD and PIGD subtypes represent different stages in the progression of PD, consistent with the arguments of Nutt (2016).

The results of this chapter provide evidence against the suitability of applying the TD and PIGD subtypes in PD. If these classifications simply represent differences in disease duration between individuals, and not the presence of different disease subtypes, it is inaccurate to use them in PD subtyping. A likely source of this inaccuracy is the approach used to quantify tremor and postural symptom severity. As this approach has not been methodologically validated, it is not clear how accurately these symptom scores represent the severity of the symptoms themselves.

3 Chapter 3 – Limitations of Current Methodologies

3.1 Structure of the MDS-UPDRS

The most established and widely used measure of PD motor symptoms is the Unified Parkinson's Disease Rating Scale (UPDRS; Goetz et al., 2008). The UPDRS was first developed in 1987 and then revised in 2008 by the Movement Disorders Society (MDS-UPDRS; Fahn et al., 1987; Goetz et al., 2008). The goal of the MDS-UPDRS is to provide a comprehensive clinical assessment of PD severity. To this end, the MDS-UPDRS is comprised of four distinct sections targeted at different dimensions of PD symptom presentation. The first, 'Non-Motor Experiences of Daily Living', is a 13-item clinician-administered questionnaire assessing the individual's perceived level of non-motor symptomology (e.g., anxious mood, hallucinations, and sleep disturbances). The second, 'Motor Experiences of Daily Living', is a 13-item self-report questionnaire assessing the individual's perceived level of motor symptomology (e.g., freezing, tremors, and speech difficulty). The third, 'Motor Examination', is a clinician-administered physical assessment rating the level of motor symptom impairment in 33 different movements (e.g., amplitude of rest tremor, gait difficulty, and joint rigidity). The fourth, 'Motor Complications', is a 6-item clinician-administered questionnaire assessing the individual's perceived level of treatment-related motor symptomology (e.g., dyskinesia, dystonia, and motor symptom fluctuations).

The 2008 paper describing MDS-UPDRS Motor Examination (Part 3) included a confirmatory factor analysis (CFA) to describe which of the 33 movements were measuring which PD motor symptom. The MDS-UPDRS was found to follow a 7-factor structure: Axial/Postural, Rest Tremor, Rigidity, Right-side Akinesia, Left-side Akinesia, Kinetic Tremor, and Lower-body Akinesia (Goetz et al., 2008). This structure has since been replicated with the Italian and Spanish versions of the MDS-UPDRS (Antonini et al., 2013; Martinez-Martin et al., 2013).

3.2 Motor Symptom Subscales

While these CFA results may give an indication of which MDS-UPDRS items are representative of which motor symptom, they do not provide a means of treating those combinations of items as actual subscales. That is, there are currently no established means of deriving scores to represent the severity of each of these seven motor symptoms. As such, there is limited consistency in the measurement of motor symptoms in PD research. Studies that purport to measure the same motor symptom often use different combinations of MDS-UPDRS items (Eggers et al., 2011; Schiess et al., 2000). Researchers typically select MDS-UPDRS items that appear most likely to measure the motor symptom of interest, but do not provide evidence to validate their selections.

Using distinct subscales to measure motor symptoms assumes that the motor symptoms measured by one subscale do not influence the measurement of another. For example, this would be the assumption that items assessing gait or balance are not affected by rigidity in the lower legs. This assumption is, however, unlikely to be accurate. Bartolic, Pirtosek, Rozman, and Ribaric (2005) reported that postural instability scores were reduced after improvements in rigidity following the use of dopamine agonists, demonstrating that the measurement of one motor symptom could be impacted by the severity of another unrelated symptom. Thus, the strength of association between a given motor score and another outcome could be over- or under-estimated, depending on the extent to which that motor symptom's measurement was affected by the severity of a different motor symptom.

The statistical approaches used by previous explorations of the MDS-UPDRS measurement structure have not been able to account for this possibility. To date, all studies have used confirmatory factor analysis (CFA). What has not yet been considered are the assumptions that CFA make about the relationships between MDS-UPDRS items and the latent motor symptom factors. That is, how the analysis conceptualises the relationship

between PD motor symptom severity and performance on MDS-UPDRS Part 3 items. CFA is underpinned by an independent latent construct model, in which it is assumed that the score on a given MDS-UPDRS item is affected solely by the latent motor symptom factor specified in the researchers' model (i.e., there are no cross-loadings unless specified). In effect, CFA does not consider the influence of other motor symptoms unless they are explicitly specified. For example, under the MDS-UPDRS 7-factor model, the severity score on Item 9 (rising from a chair) is specified as being influenced by the axial/postural motor symptom factor alone. However, it is plausible that performance on this task could be affected by an individual's level of rigidity in their knees and hips (as assessed in Item 3).

The analysis also assumes that the portions of MDS-UPDRS Part 3 item scores not influenced by their specified motor symptom are unrelated (i.e., no residual covariances unless specified). For example, both Items 8 (leg agility) and 10 (gait) require movements of the legs. Although performance on these items, primarily, influenced by their respective motor symptom factors (lower akinesia and axial/postural), both items are also frequently influenced by the general severity of symptoms in the legs. If the analysis is unable to account for this covariation, performance on a given MDS-UPDRS item may falsely appear to be more/less affected by its respective motor symptom. As such, the independent clusters model may be inadequate for assessments of the structure of the MDS-UPDRS in a manner that is congruent with how the disease would present to a clinician.

Although these assumptions may form an unrealistic view of PD, they are statistically required for the analysis to be performed. This is not to say that these assumptions, and the analyses that require them, are inappropriate in all situations. This is just to say that, given the complexity in presentation of PD, accordingly complex methods are needed to accurately model this. Unfortunately, when attempting to implement these more complex methods, the observed data often do not provide enough information to reliably estimate all of the

parameters in a given model (Brown, 2014). When this occurs, the model is described as being ‘not identified’ (Brown, 2014). To address this, many parameters are often fixed to zero, whereby the model assumes there is no relationship between given constructs (Muthen & Asparouhov, 2012). This assumption is restrictive and typically not representative of the constructs under study (e.g., the assumption of no cross-loadings in a factor analysis). One viable alternative to this was recently proposed by Muthen and Asparouhov (2012): Bayesian structural equation modelling (BSEM).

The discussion of the relative merits of Bayesian and Frequentist statistics is beyond the scope of this thesis, but the interested reader is referred to van de Schoot et al. (2014) and Zyphur and Oswald (2013). The salient difference for this thesis between Frequentist and Bayesian statistics is the way in which model parameters (e.g., factor loadings or covariances) are estimated. In the traditional frequentist framework, model parameters are estimated by selecting the values that make the observed data the most likely (hence the name: ‘maximum-likelihood’; Zyphur & Oswald, 2013). In contrast, the Bayesian framework estimates parameters by selecting values that are based on the observed data *as well as* prior information or hypotheses (Zyphur & Oswald, 2013). Frequentist estimation asks ‘given this parameter value, how likely are these data?’, whereas Bayesian estimation asks ‘given these data *and* what we already know, what parameter value is most likely?’

In the Bayesian framework, prior knowledge (commonly referred to as ‘priors’) is used to ‘weight’ the observed data and estimation process. The influence of this prior information on the resulting estimates is referred to as its level of ‘informativeness’ (Muthen & Asparouhov, 2012). Prior knowledge that is uninformative (or ‘diffuse’) generally has little impact on the estimation process, but this is not always the case (McNeish, 2016). Completely uninformative priors can skew the results in smaller samples by stating that impossibly large or small values are equally as likely as possible values (McNeish, 2016). To

avoid this, ‘weakly-informative’ priors are recommended; which state that all possible values are equally as likely but impossibly large/small values are not at all likely (van Dongen, 2006). Bayesian estimation with uninformative or weakly-informative priors produces results approximately equal to frequentist estimation (van de Schoot et al., 2014). Because of this, Bayesian estimation has been proposed as a means of estimating models that cannot be constructed in the frequentist framework (Muthen & Asparouhov, 2012)

As Bayesian estimation introduces prior knowledge into the model, we can estimate previously ‘unidentified’ parameters by including strong prior information that these parameters *should* be zero (Muthen & Asparouhov, 2012). As these priors are strongly informative, and therefore highly influential, the resulting parameters will be approximately zero (and the model will be identified). However, as parameter values are estimated by weighting the observed data with prior knowledge, if the observed relationships in the data are strongly different from zero then the prior knowledge will have little effect, and the parameter value will be different from zero. This effectively allows us to ‘fix’ the parameter value to zero (for identification) but allows the data to override this if it is not representative of what has been observed. Therefore, by estimating the CFA in a Bayesian framework the assumptions of no residual covariances and no cross-loadings can be avoided. Because these assumptions are no longer statistically required, the analysis can also explicitly test whether these assumptions are appropriate for the measure under study. Further, the identification of residual covariances and cross-loadings can be done in an exploratory manner, rather than evaluating a model post-hoc to identify which parameters to ‘free’. This is the basis for BSEM (Asparouhov & Muthén, 2011). Again, while this approach is suitable in most situations, given that the MDS-UPDRS comprises 33 items and at least 7 latent motor symptoms, iteratively freeing parameters could be inefficient and lead to the selection of an incorrect model.

A Bayesian framework is uniquely suited to analyses involving the MDS-UPDRS. The MDS-UPDRS rates motor symptom severity on a 5-point scale from 0 representing 'normal functioning' to 4 'severe impairment'. Data rated by category can be problematic to analyse. Categorical data that are treated as continuous can strongly bias results and weaken the reliability of findings (Mîndrilă, 2010). The correct treatment of categorical data with latent variables in standard maximum-likelihood (frequentist) estimation requires a process called numerical integration, which is not computationally feasible with more than four latent variables (Asparouhov & Muthén, 2012). The standard alternative is a 'limited information' estimator such as weighted least squares, which reduces the amount of information in the data to be analysed so that the model is feasible (Mîndrilă, 2010). However, these estimators are not as accurate as maximum-likelihood (especially in small samples) and rely on pairwise deletion to handle missing data. By employing a Bayesian estimator, latent variable analyses use all of the available information in the data, without pairwise deletion of missing data, and offer the most power in smaller samples of current estimators (Asparouhov & Muthén, 2012).

Thus, Bayesian analyses can explore the measurement of motor symptoms in PD in a manner that accounts for the real-world complexity in the presentation of the disease. However, Bayesian analyses have yet to be used in PD research.

3.3 Chapter 3 – Summary

This chapter discussed the lack of established and methodologically validated subscales or means of quantifying individual Parkinsonian motor symptoms using the MDS-UPDRS. Given that PD subtyping studies are not consistent in how they measure symptom severity, it is perhaps unsurprising that their conclusions are similarly inconsistent. While previous studies have explored different measurement structures of the MDS-UPDRS, the statistical assumptions of their methods (EFA/CFA) were not well suited to the complexity of PD motor symptom presentation. Despite this limitation, these assumptions are unavoidable in a traditional (frequentist) statistical framework.

A more substantively appropriate method for exploring the measurement structure of the MDS-UPDRS is available in a Bayesian framework. The use of informative priors allows the analysis to circumvent these substantively restrictive assumptions. The Bayesian framework also brings several computational benefits, foremost of which is the ability to correctly treat the MDS-UPDRS items as ordered categorical data while also being able to account for missingness.

4 Chapter 4 – Measuring Parkinsonian Motor Symptoms

4.1 Introduction

When measuring motor severity in PD, motor symptoms are typically treated as independent, whereby the severity of one does not influence the measurement of another. However, there is evidence to suggest that this independence may not be the case (Bartolic et al., 2005). Given their status as indicators of disease severity, it is important to determine whether the measurement of some motor symptoms is influenced by the presence and severity of other motor symptoms in PD.

The discussed limitations in current approaches to PD motor symptom measurement indicate that a more methodologically robust model is needed. Unfortunately, while the use of Bayesian structural equation modelling (BSEM) can circumvent the assumptions regarding residual covariances and cross-loadings, there are two other (method agnostic) issues that have yet to be addressed: general symptom severity, and symptom laterality.

4.1.1 Accounting for General Symptom Severity

General symptom severity refers to the fact that PD symptom severity typically increases with disease duration (hence its designation as a ‘progressive’ disease; Jankovic, 2008). In a given sample of individuals, each is likely to be in a slightly different stage of the disease and each is likely to have a slightly different level of overall severity as a result. As such, the MDS-UPDRS rated severity for an individual will likely be a function of both the specific motor symptom under study and that individual’s general disease severity. A given analysis needs to be able to separate the variation in severity between scores that occurs as the influence of different motor symptoms, from that which occurs due to different levels of disease duration/progression. This can be accounted for by using a bifactor approach, which statistically separates the variance shared by all items (motor symptoms) from the variance that is specific to an individual item (Reise, Morizot, & Hays, 2007). However, it is only

within the past few years that bifactor measurement models have begun to be applied in PD research (Johnson et al., 2016; Lindstrom, Wyller, Halvorsen, Hartberg, & Lundqvist, 2017; Pushpanathan et al., 2018). While a full bifactor model has not been tested with the MDS-UPDRS, a conceptually similar approach has been used with the original UPDRS to account for symptom laterality effects (Stochl, Boomsma, Ruzicka, Brozova, & Blahus, 2008).

4.1.2 Accounting for Symptom Laterality

Symptom laterality refers to the asymmetry of symptom presentation in PD (Djaldetti, Ziv, & Melamed, 2006). Some individuals will have a 'dominant' side of severity or onset which remains more severe throughout the disease (Kim et al., 2014). Analyses need to account for the fact that ratings may vary between MDS-UPDRS items not only because of different motor symptoms but also because of different symptom lateralities. As with the effects of general severity, this can be addressed by adding latent factors that influence the items assessing symptoms in each laterality (Stochl et al., 2008). This has the effect of isolating the variance in MDS-UPDRS item ratings due to laterality from the variance due to different motor symptoms; allowing the effects of each motor symptom to be more accurately estimated. Stochl et al. (2008) identified that adding these latent factors for laterality produced the best-fitting model of the original (1987) UPDRS. Despite this, no study to date has accounted for symptom laterality when assessing the measurement structure of the MDS-UPDRS.

4.1.3 Avoiding Overfitting

While adding additional latent factors can make a model more accurate in its description of the observed data (MDS-UPDRS item ratings), care needs to be taken to avoid 'overfitting' the model. Overfitting occurs when a model is highly accurate in describing the currently observed data but is too specific to be applied to other individuals (Myung, 2000).

There are two main approaches used to account for overfitting: adjusted fit indices and replication in another sample.

4.1.3.1 Adjusted Fit Indices

Adjusted fit indices refer to the use of model fit statistics that assess the ability of a given model to predict the performance of an individual not in the current sample (Vehtari, Gelman, & Gabry, 2016). That is, the assessment of a model's ability to be applied to another group of individuals. For Bayesian models, a commonly-used (but not Bayesian-specific) approach is Leave One Out (LOO) cross-validation (Ross et al., 2009). The full implementation of LOO validation requires partitioning the sample into several groups and repeatedly fitting a given model, each time with a different group of the sample removed (Ross et al., 2009). While this produces the most precise result, the time taken is not feasible in practice for large samples and multiple complex models. An alternative, the LOO Estimated Log-Posterior Density (ELPD), was proposed to address this limitation (Vehtari et al., 2016). The LOO ELPD statistic allows for the approximation of how well a given model would perform in a different sample, without requiring repeated estimation (Vehtari et al., 2016). However, it should be noted that the ELPD statistic can be biased towards more complex models (Gronau & Wagenmakers, 2019). As such, a model chosen by the ELPD should be subsequently validated by a more thorough method.

4.1.3.2 Invariance Testing

Though statistical methods can approximate the accuracy of a model across samples, the most accurate method is to sample (at least) two groups and test whether the model is invariant between them (van de Schoot, Lugtig, & Hox, 2012). Model invariance testing is the assessment of whether the measurement structure of a scale is the same in multiple groups (van de Schoot et al., 2012). Unless the measurement structure of the MDS-UPDRS is established as being invariant across samples, it is not appropriate to compare the ratings of

individuals between them (van de Schoot et al., 2013). A lack of invariance means that the MDS-UPDRS is measuring motor symptoms differently in either group, and by extension, that the measured severity scores do not represent the severity of the same motor symptoms.

While there are increasingly strict tests of invariance (equality between groups), the minimum required for accurate comparisons of group outcomes on a scale is the invariance of factor loadings and thresholds (van de Schoot et al., 2013). For the MDS-UPDRS, factor loading invariance is the assessment of whether motor symptom severity affects ratings on the MDS-UPDRS in the same way in different groups of individuals. Threshold invariance is the assessment of whether the same level of observed symptom severity was associated with the same category rating (e.g. 'Normal' or 'Severe') on MDS-UPDRS items. Despite several studies using scores from the MDS-UPDRS to compare different samples, an invariant measurement structure has yet to be determined.

The aim of the present study was to (a) develop, and (b) validate, a model of motor symptom measurement in PD using the MDS-UPDRS. Model development involved testing a sequence of models, each addressing different limitations in current approaches, and statistically assessing which maintained the best balance of accuracy and generalisability to another sample. The best model of the MDS-UPDRS was then validated by testing whether its measurement was the same (invariant) in a separate sample.

4.2 Methods

4.2.1 Participants

The first (model development) sample comprised 248 participants from the ongoing ParkC study at Curtin University (ethics committee approval number: HR158/2013; Loftus et al., 2015). Inclusion required a diagnosis of PD by a geriatrician/neurologist in accordance with the United Kingdom Parkinson's disease Brain Bank Criteria (Hughes, Daniel, Kilford,

& Lees, 1992). The ParkC sample comprised 169 males (68.14%) and 79 females (31.86%), with a mean age of 65.67 years (*SD* = 14.24), mean disease duration of 69.82 months (*SD* = 62.60), and mean Levodopa equivalent dose of 627.61mg (SD = 497.26; Tomlinson et al., 2010). All individuals in the ParkC sample were assessed in the 'on' state of their medication.

The second (model validation) sample comprised 423 de novo (unmedicated) individuals, diagnosed within two years prior to assessment, from the Parkinson's Progression Marker Initiative (PPMI; Marek et al., 2011). The PPMI sample comprised 277 males (65.48%) and 246 females (34.52%), with a mean age of 61.69 years (*SD* = 9.72).

4.2.2 Measures

Motor symptoms were assessed using Part 3 of the MDS-UPDRS (Goetz et al., 2008). Part 3 comprises 33 items assessing the motor symptoms of PD and is administered by a trained researcher or clinician. Motor symptoms are rated on a 5-point scale from 0 (Normal) to 4 (Severe).

4.2.3 Analysis

All analyses were conducted using the RStan package (version 2.19.1) in R 3.5.1 (Carpenter et al., 2017). RStan uses the No U-Turn Sampler (NUTS), a self-tuning Hamiltonian Monte Carlo algorithm which (generally) produces estimates of higher quality in fewer iterations than other Bayesian samplers (e.g., Gibbs or Random-Walk Metropolis; Carpenter et al., 2017). Analyses were estimated with four chains for 20000 iterations (80000 iterations total), with the first half of each chain discarded as 'warmup', the results presented are based on the remaining 40000 iterations. RStan also allows for the manual specification of 'quality parameters' for the analysis that improve accuracy at the cost of a longer estimation time. For all analyses, the average acceptance ratio ('adapt_delta' in RStan) was set to 0.95, and the maximum treedepth ('max_treedepth') to 15.

Estimating factor models in a Bayesian framework comes with several unique requirements and difficulties. The sources and remedies are too complex to be adequately covered in this methods section and are instead provided as a series of appendices. The statistical background for constructing a factor model in a form that can be sampled in a Bayesian context is discussed in Appendix A. The difficulties unique to the Bayesian estimation of factor models, and how they were addressed in the present analysis, are discussed in Appendix B. The general structure of a Stan program is explained in Appendix C. An annotated version of the final Stan model that explains each stage of the construction/specification is covered in Appendix D. The final model is then presented in full in Appendix E. The distributions for each parameter in the final model are summarised in Appendix F.

There are two primary aspects of Bayesian model checking: convergence and fit. Convergence refers to whether the analysis has arrived at the 'correct' estimates for the model, whereas fit refers to how well the model represents the observed data. When a Bayesian (MCMC) sampler is used, the analysis estimates a set of values for all parameters in the model and then, based on those values, estimates another set, and another, and so on (Kruschke, 2014). This process is referred to as a 'chain' of estimates. Convergence occurs when these chains are all estimating sufficiently similar values. This is assessed using the Potential Scale Reduction Factor (PSR). The PSR is a measure of the similarity in estimates; comparing the variation of estimates within chains to the variation between chains. A PSR $\leq$ 1.05 indicates good convergence (Kruschke, 2014). All analyses reported a PSR $\leq$ 1.01 for all parameters.

Fit was determined as the balance of two statistics: the deviance information criteria (DIC), and the previously mentioned LOO ELPD. The DIC can be thought of as a Bayesian analogue of the Akaike Information Criterion (AIC), with which readers may be more

familiar. Information criteria such as DIC and AIC are used for comparing the fit of competing models, rather than as an indicator of the quality of a single model. Lower values of the DIC statistic indicate a better fitting model. The DIC fit statistic can, however, be overly sensitive to different model parameterisations (Plummer, 2008) and should be supplemented with a more robust fit statistic, such as the LOO ELPD.

The LOO ELPD is interpreted in the same fashion as the log-likelihood from a traditional (frequentist) analysis (i.e., larger values indicate better fit). It is also possible to statistically compare the fit of two models by constructing confidence intervals for their LOO ELPD difference. If the confidence intervals for the difference in fit contain zero, this implies a possibility that the 'true' difference may in fact be zero (i.e., that there is no difference in fit between the two). Where standard frequentist confidence intervals are constructed using $\pm$ 1.96*SE, the derived standard errors for the LOO ELPD difference can be underestimated, and so a more stringent interval is recommended. The authors of the fit statistic recommend confidence intervals of $\pm$ 3*SE (Vehtari, 2017).

Bayesian analyses are also interpreted somewhat differently (Kruschke & Liddell, 2018). Rather than a single estimate for a parameter, Bayesian analyses return a distribution of possible values. The mean of this distribution is then used as a single representative estimate (Kruschke & Liddell, 2018). This is interpreted in combination with the parameter's 95% credibility intervals (the 2.5% and 97.5% quartiles of the estimated posterior distribution), which show the range within which the 'true' population parameter is 95% likely to fall. If the credibility intervals do not contain 0, this adds confidence to any inference of whether a relationship is present (Kruschke & Liddell, 2018).

4.2.3.1 Model Priors

Priors for the normal distribution will be presented with the notation $N(M, V)$, where M refers the mean of the distribution, and V refers to the variance. Item thresholds and intended factor loadings were given weakly-informative priors of $N(0, 25)$, stating that 99% of values (3 standard deviations) should fall in the range [-15.00, 15.00]. Cross-loadings were assigned informative priors of $N(0, 0.01)$, stating that 99% of values should fall in the range [-0.3, 0.3]. Correlations/covariances were given an LKJ prior (named for the authors that proposed the distribution; Lewandowski, Kurowicka, & Joe, 2009). The LKJ prior distribution is controlled by a single parameter η, which determines the strength of prior information about correlations between items/constructs. When $\eta = 1$, the prior is uniform and uninformative (i.e., all possible correlation values are considered equally likely). As the value of η increases, correlations closer to 0 are viewed as increasingly more likely. Correlations between the motor symptom factors were given an LKJ prior with $\eta = 1$, giving equal support to all possible values. Residual covariances between MDS-UPDRS items were given an LKJ prior with $\eta = 20$, giving strong prior information of zero residual covariance.

4.2.3.2 Model Development

To develop a model of motor symptom measurement in PD, a sequence of different measurement structures was fitted to the ParkC MDS-UPDRS data:

(1) CFA: The original 7-Factor model proposed with the revised MDS-UPDRS. Hypothesises assessment of 7 motor symptoms: Axial, Rest Tremor, Rigidity, Right Akinesia, Left Akinesia, Lower Akinesia, and Kinetic Tremor. This model contains no cross-loadings and no residual covariances. This is the same model that was presented in the original publication of the MDS-UPDRS (Goetz et al., 2008)

(2) BSEM with Approximate Zero Cross-Loadings: The same model as the previous, but with the addition of cross-loadings to all MDS-UPDRS items. The added cross-loadings assess the impact of motor symptoms on MDS-UPDRS items that were not originally hypothesised.

(3) BSEM with Approximate Zero Cross-Loadings and Approximate Zero Residual Covariances: Residual covariances between the observed MDS-UPDRS items are added. The added covariances assess the presence of relationships between item ratings that cannot be explained by the model.

(4) BSEM Bifactor with Approximate Zero Cross-Loadings and Approximate Zero Residual Covariances: A bifactor is added to the model to control for the effects of general disease severity.

(5) BSEM Lateral Bifactor (Stochl et al., 2008) with Approximate Zero Cross-Loadings and Residual Covariances: The general severity bifactor is replaced with two laterality factors (i.e., one factor affecting right-sided MDS-UPDRS items, and one factor affecting left-sided MDS-UPDRS items).

(6) BSEM Lateral Bifactor (Stochl et al., 2008) with Single Akinesia Factor. As symptom laterality was already being accounted for by the two bifactors, the left- and right-sided akinesia factors were collapsed into a single ‘akinesia’ factor

4.2.3.3 Model Validation

The validation (invariance) testing process involves freely estimating the model in both groups (ParkC & PPMI) and then increasingly constraining parts of the model to be equal in both. For the present analysis, it was of interest to determine whether both the thresholds and factor loadings were invariant across groups. If the factor loadings are invariant, this indicates that the different motor symptoms affect the severity ratings on MDS-

UPDRS items in the same way in both groups. If the thresholds are invariant in both groups, this indicates that each category rating represents the same level of observed severity in both groups. Both threshold and loading invariance is required for comparing groups on a given outcome.

4.3 Results

4.3.1 Model Development

When comparing the 7-Factor CFA to the 7-Factor BSEM, there was a marked improvement in fit when cross-loadings with informative priors were added to the MDS-UPDRS 7-Factor model (see Table 1). Despite a wide range in the confidence intervals around the difference in LOO ELPD model fit, they do not cross zero. This suggests that the MDS-UPDRS items are not solely affected by the motor symptoms that they were designed to measure. That is, the rating on a given MDS-UPDRS item appears to be affected by the severity of more than one motor symptom. While we cannot precisely estimate the increase in accuracy when allowing for this, we can be confident that the improvement is larger than zero.

Table 3
BSEM Model Fits and Model Fit Differences

Model	DIC	ELPD (LOO)	ELPD Difference [95%CI]
7-Factor CFA	14177.80	-6081.88	
7-Factor BSEM	14340.89	-6001.96	79.92 [60.92, 98.91]
7-Factor BSEM (Residual Covariances)	14612.85	-5899.80	102.17 [86.08, 118.25]
7-Factor Bifactor	13743.87	-5714.20	185.597 [132.54, 238.65]
7-Factor Lateral Bifactor	14492.83	-5663.85	50.35 [25.15, 75.56]
6-Factor Lateral Bifactor	14217.88	-5669.09	-5.24 [-18.78, 8.31]

Similar improvements were seen when allowing for residual covariances between the MDS-UPDRS items; there was a marked improvement in fit with large confidence intervals that do not cross zero. This would suggest that the severity ratings between items were related in some way that the model could not account for. The addition of a bifactor to account for individual differences in general severity improved the fit of the model even further, but to a smaller degree than the previous model. Replacing the bifactor with two factors to account for individual differences in motor laterality saw smaller again improvements. Collapsing the two akinesia factors ("right-" and "left-sided" akinesia) into a single akinesia factor showed a very small decrease in the model's fit to the data, but with confidence intervals for the difference that crossed zero. This indicates that we cannot be confident of a difference in model fit between the 6- and 7-Factor models with bifactors for laterality. As such, the more parsimonious (6-Factor lateral) model is preferred. Table 4 presents the standardised loadings for all parameters whose credibility intervals did not contain zero, with the final model diagrammed in Figure 11.

Table 4

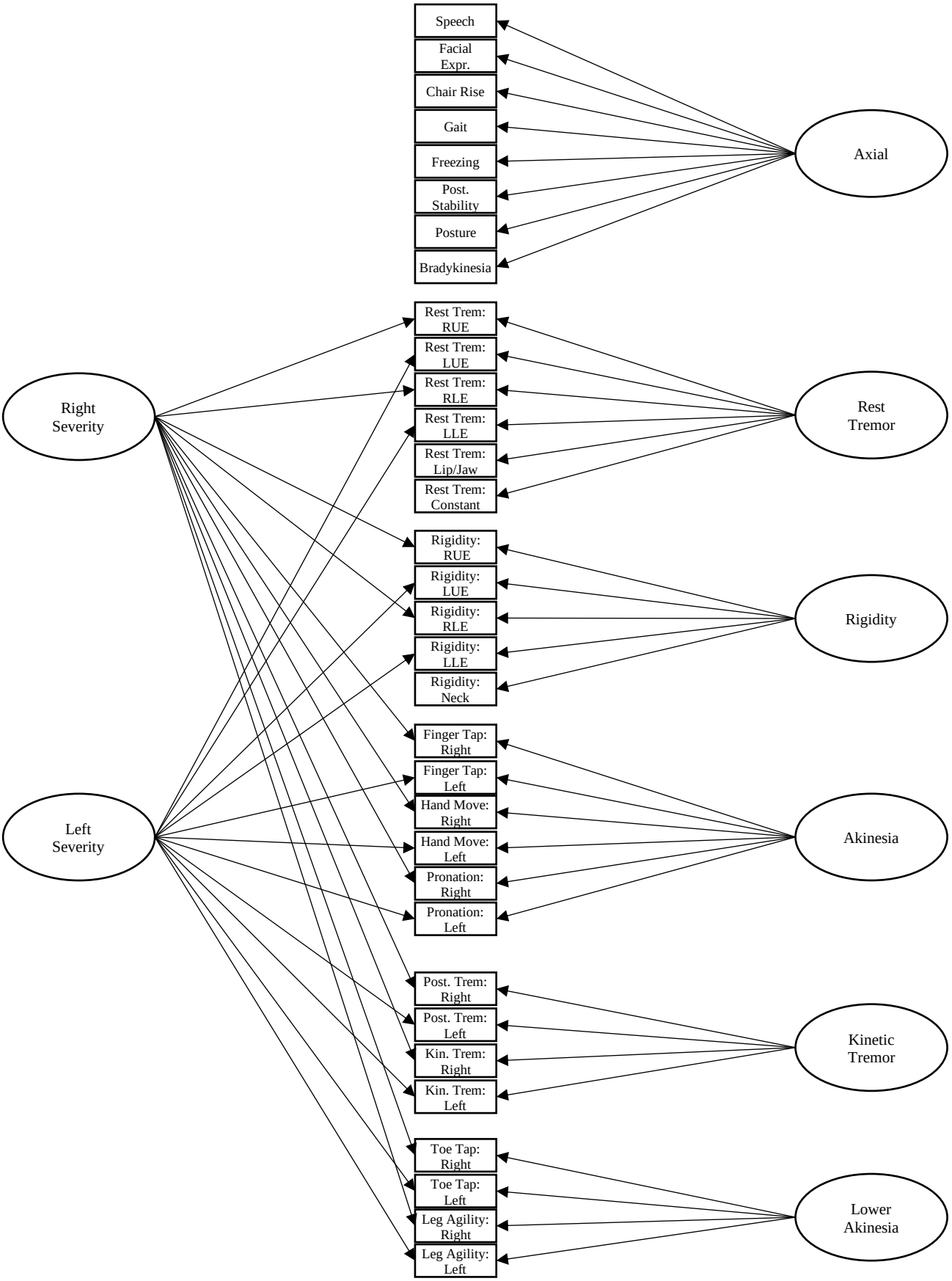
Standardised Factor Loadings and 95% Credibility Intervals for Lateral Bifactor BSEM

MDS-UPDRS Item	R	L	F1	F2	F3	F4	F5	F6
F1. Axial								
Speech			.91 [.78, 1.03]					
Facial Expression			.83 [.67, .98]					
Chair Rise			.90 [.77, 1.01]					
Gait			.94 [.83, 1.03]					
Freeze Gait			.45 [.08, .66]					
Post. Stability			.83 [.66, .98]					
Posture			.82 [.63, .98]					
Bradykinesia			.93 [.84, .98]					
F2. Rest Tremor								
Rest Trem. RUE	.71 [.56, .88]			.41 [.14, .69]				
Rest Trem. LUE		.50 [.35, .64]		.66 [.47, .82]				
Rest Trem. RLE	.60 [.30, .86]			.42 [.07, .76]				
Rest Trem. LLE		.42 [.12, .73]		.58 [.20, .85]				
Rest Trem. Jaw				.73 [.46, .92]				
Rest Trem. Const.				.98 [.94, 1.02]				
F3. Rigidity								
Rig. Neck					.54 [.32, .73]			
Rig. RUE	.38 [.18, .58]				.76 [.61, .90]			
Rig. LUE		.47 [.30, .68]			.76 [.62, .88]			
Rig. RLE	.11 [.01, .27]				.97 [.90, 1.02]			
Rig. LLE		.28 [.12, .43]			.92 [.84, .99]			
F4. Akinesia								
Finger Tap R	.33 [.18, .58]					.79 [.65, .91]		
Finger Tap L		.61 [.47, .74]				.69 [.54, .84]		
Hand Move. R	.33 [.11, 55]					.86 [.75, .96]		
Hand Move. L		.62 [.48, .75]				.74 [.58, .88]		
Supination R	.36 [.14, .57]					.84 [.71, .94]		
Supination L		.63 [.49, .75]				.69 [.52, .84]		
F5. Lower Akinesia								
Toe Tap. R	.48 [.29, .66]						.73 [.58, .86]	
Toe Tap. L		.52 [.35, .68]					.66 [.48, .81]	
Leg Agility R	.26 [.07, .44]						.90 [.80, .98]	
Leg Agility L		.42 [.26, .56]					.83 [.72, .93]	
F6. Kinetic Tremor								
Post. Trem. R	.62 [.30, .83]							.47 [.23, .76]
Post. Trem. L		.56 [.27, .77]						.55 [.30, .78]
Kin. Trem. R	.39 [.09, .64]							.67 [.43, .89]
Kin. Trem. L		.45 [.17, .67]						.66 [.44, .86]

Note. Loadings with Credibility Intervals Containing 0 are Suppressed;

Figure 11

Lateral Bifactor BSEM – Salient Loadings and Residual Covariances



Note. Loadings with Credibility Intervals Containing 0 are Suppressed;

4.3.2 Model Validation

Once the best-fitting measurement model of the MDS-UPDRS had been determined, it was validated using data from the PPMI cohort. The 6-Factor lateral bifactor model was first freely estimated in both the ParkC and PPMI cohorts. When constraining the factor loadings to be equal between the two groups, there was no marked change in fit (confidence intervals containing zero; see Table 5). This indicates that the estimated factor loadings (effects of motor symptoms) describe both the ParkC and PPMI cohorts equally well. However, the model fit clearly worsened (with both confidence intervals excluding zero) when constraining the item thresholds to be equal. This would suggest that the levels of observed symptom severity represented by each rating category were different between the ParkC and PPMI cohorts.

Like the restrictive assumptions of zero cross-loadings and zero residual covariances discussed earlier, the assumption of absolute equality in item thresholds may also be unrealistic. Given that the two samples were assessed by different clinicians, each with different levels of experience and training, it is reasonable that each would rate motor symptoms in a different way. This is not to say that some clinicians were incorrectly rating individuals, but that each clinician would likely have a slightly different internal model of the level of severity required for each category rating. Beyond this, each clinician would likely have a different level of skill in identifying the subtler presentations of symptom severity (e.g., amount of hesitancy in the finger-tapping task). As such, rather than testing for exact equality of thresholds (exact invariance) we can instead test for *approximate* equality (approximate invariance).

van de Schoot et al. (2013) identified that a prior of $N(0, 0.25)$ for the difference in thresholds resulted in no bias when there were only small differences between the groups. By using this strongly-informative prior, we can test the assumption that the thresholds in the

PPMI cohort are within 1.0 (2 SDs) of their respective thresholds in the ParkC cohort. Re-fitting the model with approximately invariant item thresholds fit the data far better than the model with exact invariance of item thresholds. The model fit was also no longer markedly different than that of the model with all loadings and thresholds freed. The final step was to fit the model with invariant factor loadings and approximately invariant item thresholds. This model showed no marked difference in model fit from the model with freely estimated loadings and item thresholds (See Table 5). Furthermore, the credibility intervals for the estimated differences in thresholds between the cohorts all crossed zero. Together, these results indicate that a 6-Factor lateral bifactor structure of the MDS-UPDRS is present in both the ParkC and PPMI cohorts.

Table 5
Invariance Testing Model Fits and Model Fit Differences (Compared to 'All Free' Model)

Model	DIC	ELPD (LOO)	ELPD Difference [95%CI]
All Free	33878.50	-13844.05	
Loadings Invariant	33893.13	-13881.32	16.76 [-87.55, 13.00]
Thresholds Invariant	34246.41	-14039.19	23.27 [-264.96, -125.33]
Loadings Invariant, Thresholds App. Invariant	33644.39	-13887.47	17.61 [-96.04, 9.62]

However, beyond the statistical evidence for the 6-Factor lateral bifactor structure, it is also useful to explore, on an individual basis, how this more complex model can provide more sensitive measurements of motor severity. The Axial symptom severity for three individuals, estimated under the original MDS 7-Factor CFA and the proposed 6-Factor lateral bifactor structure, are presented in Table 6. Because of how the models were specified, these estimates are standardised, and so a 1-point difference represents a 1 standard deviation difference. It is clear that the estimated Axial severity is over 1SD higher under the CFA

model than under the lateral bifactor model. This difference is due to the lateral bifactor model being able to account for the differences in severity due to laterality, as well as the residual covariances between items. In effect, the simpler 7-factor CFA model results in a biased estimate of an individual's motor symptom severity. While more complex, the 6-factor lateral bifactor model can provide a more precise measure of an individual's motor symptom severity.

Table 6
Estimated Axial Severity Under 7-Factor CFA and 6-Factor Lateral Bifactor Model

Individual	7-Factor CFA	6-Factor Lateral Bifactor
Individual 1	3.55	2.54
Individual 2	3.29	2.33
Individual 3	3.61	2.72

4.4 Discussion

The present study aimed to develop and validate a model of motor symptom measurement using the MDS-UPDRS in two community cohorts of individuals with PD. The analyses identified six motor symptoms under assessment: Axial, Rest Tremor, Rigidity, Akinesia, Lower Akinesia, and Kinetic Tremor. The analyses further identified that these motor symptoms differed by laterality (i.e., left- and right-sided). This 8-factor structure was replicated in a separate sample of individuals assessed as part of the Parkinson's Progression Marker Initiative (Marek et al., 2011).

Overall, this demonstrates that MDS-UPDRS item scores are influenced by unrelated items (evidenced as residual covariances) and by laterality (demonstrated by the improved fit of the lateral bifactor model). As such, a method of scoring that accounts for these is

required. This is straightforward for researchers who are able to fit the previous factor model and generate factor scores for use in further analyses. For the clinician assessing a single individual, this is not feasible (as a factor analysis requires more than one individual). One solution is to estimate 'coarse' factor scores, by weighting each MDS-UPDRS item score using its loading on the motor factor and then summing. A full table of loadings and an excel file which calculates these scores from an individual's MDS-UPDRS item scores are provided in Supplementary Materials A and B, respectively.

It should be emphasised that the 'coarse' symptom scores are unsuitable for research/correlational purposes. Although the scores may accurately assess their respective symptoms, the correlations between the 'coarse' estimates are likely to be highly inflated (Grice, 2001). It is for this reason that researchers are encouraged to fit the aforementioned factor model and generate factor scores, as the process has been refined to reduce the overestimation of correlations (DiStefano, Zhu, & Mindrila, 2009).

These results have implications for the assessment of motor severity and laterality in PD. The residual covariances and bifactor models suggest that the measurement of a symptom (e.g., axial) is likely influenced by the severity of another (e.g., lower akinesia). The reduction in covariances and improved fit when moving from the single to the lateral bifactor models also suggests that left-sided symptoms vary in severity differently to right-sided symptoms. This is consistent with Kim et al. (2014), who identified that laterality of motor symptoms was affected by cortical thinning patterns in PD. Mean/sum scores of MDS-UPDRS items are therefore likely influenced by both the severity of unrelated items and the laterality of items themselves.

Research examining motor severity and clinical outcomes almost exclusively use sum/mean scores. This has implications for the accuracy of reported results. Given the

likelihood of measurement error being present, as demonstrated by these analyses, the risks of Type 1 and 2 errors are inflated. Using a summed/averaged score of items may not yield an accurate representation of symptom severity, as the score would likely be influenced by overall lateral severity and by the severity of other specific symptoms. This bias provides a possible account for the heterogeneity in PD subtyping results to date. If studies were not using accurate measures of motor symptoms, then it is possible that their results may not accurately reflect the motor subtypes of PD.

These findings further impact clinicians and clinical practice. Without a standard, established, method for scoring individual motor symptoms it is likely that each clinician will use a slightly different combinations of MDS-UPDRS items. This lack of consistency between practitioners could make it difficult for a new clinician to use a previous clinician's data as an indication of symptom severity, adversely impacting on the capacity to provide effective treatments.

This study is the first to examine the MDS-UPDRS using BSEM and the first to fit a bifactor measurement model to the MDS-UPDRS. Further exploration using BSEM in an independent sample is required. It should be noted that all current assessments were conducted in the 'ON' state of medications, which is common in previous studies of the UPDRS and MDS-UPDRS (Goetz et al., 2008; Stochl et al., 2008).

The present study highlights the difficulties with PD motor symptom measurement, and the impacts of symptom laterality and severity of purportedly unrelated symptoms. Mean/sum totals of MDS-UPDRS items may therefore not accurately represent the relationships between motor symptoms in PD. We recommend that researchers and clinicians estimate symptom scores based on the factor structure presented in this paper and the supplementary materials provided. Having derived a more accurate and methodologically

robust model of PD motor symptomology, the assessment of PD motor subtypes can proceed in a similar fashion.

4.5 Chapter 4 – Summary

This chapter derived, and validated, a means of measuring individual motor symptoms and motor symptom laterality. The results indicated that there were six unique motor symptoms being measured by the MDS-UPDRS: axial, resting tremor, rigidity, kinetic tremor, akinesia, and lower akinesia. The results provided further evidence that the severity of these motor symptoms differed by laterality (i.e., left- vs. right-sided).

The improved performance of the methods used in the present chapter (BSEM) indicate that quantifying Parkinsonian motor symptom severity requires a more robust approach than summing or averaging MDS-UPDRS item scores. The results indicated that the scored severity of a given motor symptom (e.g., rigidity) can be influenced by the severity of a motor symptom that is not being measured (e.g., akinesia). It is essential for both research and clinical practice to use measurements that are the best representation possible of the symptoms under study, otherwise the conclusions will themselves be inaccurate. By using the methods provided in the present chapter (and the calculator in the supplementary materials), such representative measurements are now possible.

This chapter provides a model for measuring and quantifying Parkinsonian motor symptom severity. This model can now be integrated into an analysis aiming to identify subtypes based on individual differences in motor symptom severity.

5 Chapter 5 – A New Approach to PD Subtyping

5.1 Introduction

The concept of 'grouping' individuals with Parkinson's disease (PD) by the severity of different symptoms was first demonstrated, and popularised, by Jankovic and colleagues in 1990. In their seminal paper, "Variable expression of Parkinson's disease: A baseline analysis of the DATATOP cohort", Jankovic and colleagues proposed grouping individuals with PD based upon the difference in severity between their tremor and postural symptoms (Jankovic et al., 1990). The approach had been discussed in work several years prior but had yet to gain traction in the broader clinical and research communities in PD. Jankovic et al.'s approach of comparing the magnitude of mean tremor and postural symptom scores quickly rose in popularity and continues to be applied in current research (Pelicioni et al., 2018). Despite this continued popularity, Jankovic et al.'s approach is problematic. This was demonstrated earlier in Chapter 2, where the magnitude of difference between tremor and postural symptoms was shown to increase with disease duration. However, the alternatives that have been proposed to replace Jankovic et al.'s method have similar limitations.

Whilst Jankovic et al.'s (1990) method was useful as a means of 'popularising' the field of PD heterogeneity, it also became clear that the approach was not comprehensive enough. While this approach separated individuals by severity of two motor symptoms (tremor and postural), other studies had observed individual differences in a range of other PD symptoms (Kehagia et al., 2010). However, while it was clear that individuals differed in their severity of these symptoms, it was not clear which could be used as a means of separating individuals into groups (and how). This required a more exploratory form of analysis; where a selection of clinical, cognitive, or motor assessments could be evaluated for their ability to separate individuals with PD into distinct groups.

5.1.1 K-Means Cluster Analysis

The first attempts at implementing this 'exploratory' approach to subtyping used K-means clustering (Graham & Sagar, 1999; Lundervold, Karlsen, & Reinvang, 1994). This attempts to separate individuals into *K* groups, where the mean of each group is as different as possible to the means of the others (Jain, 2010). While K-means clustering has seen success in some applications, the statistical approach of the analysis was not well suited to quantifying heterogeneity in PD. K-means clustering aims to separate individuals into groups that are clearly distinct from one another (Jain, 2010). The analysis approaches this task by using so-called 'hard' clustering. Under a 'hard' clustering approach, individuals are evaluated as belonging to a single subtype (Singh, Tiwari, & Garg, 2011). This means that if an individual is only marginally more likely to belong to one group than another, this is not considered (or presented in the output). Where individuals' symptoms differ in severity in a relatively clear manner, this is not an issue. However, when symptom heterogeneity is not so clearly distinct (as is often the case in PD), the results of the analysis are likely to be less accurate by not taking this uncertainty into account (Magidson & Vermunt, 2002b).

5.1.2 Latent Profile Analysis

As such, because the symptom groupings in PD are not so clearly distinguished, a more appropriate analysis was needed. This led to the use of Latent Profile Analysis (LPA). LPA and K-means have some core differences in how they approach grouping individuals. Where K-means identifies the single group to which an individual is most likely to belong, LPA assesses that individual's likelihood of belonging to each of the identified groups in an analysis (Berlin, Williams, & Parra, 2013). In effect, LPA explicitly models the possibility that individuals will not present in well-defined groups. Because of this additional consideration of classification uncertainty, LPA generally outperforms K-Means in comparisons of subtyping performance (Magidson & Vermunt, 2002b). An LPA-based

approach shows clear utility for use in PD, where individuals may only be mildly, rather than markedly, different from one another.

Latent profile analysis has since become the ‘de facto’ method for assessing symptom heterogeneity in PD. Studies have applied LPA to an ever-increasing range of symptoms/assessments in an attempt to develop a means of grouping individuals (Marras & Lang, 2013). However, because LPA is so easily applied to a large range (and number) of variables, it is increasingly being treated as a ‘black box’ of sorts in the PD subtyping field. An increasingly common trend is for studies to simply introduce a large combination of symptoms (e.g., genetic, cognitive, or motor) to which they apply LPA, and then discuss the different groups that are identified. There are two primary implications for this approach to subtyping. Firstly, this is not taking full advantage of how LPA can be used, or the kinds of hypotheses that can be tested. Secondly, the statistical assumptions made by LPA can be inconsistent with the disease and measurements under study.

5.1.2.1 Controlling for Measurement Error

Latent profile analysis is extremely flexible in how it can be applied to grouping individuals. It is often described as ‘model-based clustering’, because it is effectively just a latent variable model (Lubke & Luningham, 2017). As such, the clustering process of LPA can be integrated with other, more complex, models of symptoms or symptom relationships in PD (Lubke & Muthen, 2005). A key example of this application is in the concept of measurement error. When applying LPA to a set of symptom measurements, the analysis effectively treats those measurements as perfect representations of the symptoms being assessed. If this is not the case, then analyses that group individuals based on these measurements will not be accurate representations of the symptom groups that are present. Take as an example the assessments of akinesia in the MDS-UPDRS. These assessments involve the repeated movements of limbs (legs, arms, feet, and fingers) with akinesia severity

rated by the amount of hesitancy and changes in amplitude of movement. If individuals have some level of rigidity in these joints, it is likely that their ability to initiate movements will be impaired. If a subtyping analysis is not able to account for this, it may mistakenly identify a group as having high levels of akinesia, when in fact these individuals had high levels of rigidity.

5.1.2.2 Controlling for Covariates

A similarly important case is that of controlling for covariates. Because LPA is a latent variable model, it is a simple matter to control for the effects of a covariate while subtyping (Lubke & Luningham, 2017). That is, it is a simple matter to remove/separate the individual variation in motor symptoms due to differences in antiparkinsonian medication and disease duration from that due to different disease subtypes. However, this has yet to be done in PD. An approximation of this has been applied wherein the authors will enter the covariates and outcomes of interest in a linear regression and save the residuals for use in a separate analysis (van Rooden et al., 2011). There are, however, concerns with this approach.

The key problem in using the residuals from a separate regression is that the analysis is not considering the effects of either disease subtype or measurement error when evaluating the effects of the covariates. This means that the analysis may under- or over-estimate the effects of covariates when estimating these residuals, and so they would not be accurate representations of motor symptoms. To be both accurate (and precise) in separating the effects of covariates and subtypes, the analysis needs to evaluate these effects simultaneously (i.e., in the same model). By not doing this simultaneously (as in the two-step process of subtyping on residuals), the analysis is effectively introducing error from two sources. There is first the error in estimating the effects of covariates without considering the effects of motor subtypes. There is then the error in attempting to derive subtypes from biased measurements.

The appropriate inclusion of disease duration in the model also bears discussion. The results of the longitudinal analysis in Chapter 2 indicated that the TD/PIGD subtypes may represent stages of disease progression rather than distinct biological subtypes. Given this, it is important to consider the appropriate method of accounting for individual differences in disease duration. The standard approach, and that employed by van Rooden et al. (2011), is to introduce disease duration as a control variable. When introducing duration as a control variable, the subtyping analysis is then attempting to identify groupings within the sample at a particular length of disease duration. In other words, this approach is asking the question: "at the sample average disease duration, what subtypes are present?". However, a staging hypothesis of subtyping implies that the presentation of subtypes is dependent on the disease duration of the individual. If this is the case, the subtypes that are present at the sample average disease duration may not be representative of the groupings that are present at different disease durations.

5.1.2.3 The Assumption of Local Independence

However, there is a larger concern with the application of LPA: the assumption of local independence. In the context of a latent profile analysis, local independence refers to the assumption that the extracted subtypes should fully explain the observed relationships between the items under analysis (Oberski, 2016). In more practical terms, this implies that within each of the estimated subtypes there should be no relationship (or covariance) between the measured symptoms (e.g., motor or cognitive; Oberski, 2016). This is a foundational assumption, and one that is not discussed in the PD subtyping literature.

When considering a subtyping analysis using only a small group of variables (e.g. only measures of motor symptom severity), this is not of much concern. After all, the intent of a subtyping analysis is to explain why measurements appear related (because of different populations). However, this is a large concern when studies are conducting latent profile

analysis on measures that are highly likely to be related, regardless of cluster or even PD. Take for example the LPA conducted by van Rooden et al. (2011) which included measures of sleep quality and anxiety in PD. While it is substantively reasonable to suggest that the relationship between anxiety and sleep quality may be different between PD subtypes, this is not what LPA hypothesises. Rather, LPA suggests that sleep quality and anxiety only appear related in PD because there are different subtypes with different levels of each. In other words, after accounting for the different levels of anxiety and sleep quality in each subtype, there should be no relationship between the two. This is not substantively reasonable, given the well-established relationship between anxiety and sleep quality in multiple clinical and non-clinical populations (Gregory et al., 2011; Ramsawh, Stein, Belik, Jacobi, & Sareen, 2009).

Despite the substantive issues, it is becoming increasingly common for PD subtyping studies to include clinical measures that are likely to be related beyond what would be expected in PD. When this is the case, a latent profile analysis is primarily tasked with separating individuals into groups where these variables are not related, when that is not the aim of the research or even realistic. By allowing relationships between these measurements while subtyping, the analysis would instead attempt to find groups where these relationships are different – which is generally more substantively interesting and realistic. However, this is restricted by the same limitation as those in the case of residual covariances and cross-loadings: model identification (Asparouhov & Muthén, 2011). There is not enough information provided in the observed data to be able to estimate both the subtypes, as well as the relationships within those subtypes (and so such a model would be not identified). In a Bayesian framework however, we can introduce this required (prior) information.

5.1.3 Bayesian Implementations of Subtyping

Bayesian subtyping analyses have seen a brief introduction to PD through the work of White et al. (2012). The authors used PD subtyping as a case study in their argument for the use of Bayesian estimation methods in LPA. However, as the subtyping itself was not the primary aim of the study, it was not conducted in great depth. White et al. (2012) used only the observed MDS-UPDRS items for subtyping (i.e., did not control for measurement error) and did not address the assumption of local independence.

Relaxing the assumption of local independence in a Bayesian latent profile analysis is conducted in the same manner as that used for residual covariances in the previous chapter. By introducing strong prior information that there should be no relationships between variables within each subtype, the model will have sufficient unique information to be identified (Asparouhov & Muthén, 2010). However, this will also allow for relationships to exist (despite the prior information) if they are present. This is because Bayesian estimation is a combination of both the introduced prior expectation and the observed relationships. If the observed data differs from the prior expectation (of no relationship), it will effectively 'override' the prior and the estimate will reflect the observed relationship (Muthen & Asparouhov, 2012).

In sum, the assumption of local independence in latent profile analysis is a strong concern with current subtyping analyses, effectively putting the purpose of the analysis at odds with its stated aims. While the 'soft clustering' approach of latent profile analysis is well-suited to the non-distinct heterogeneity in PD, local independence introduces a dissonance between the statistical and research hypotheses that needs to be addressed. By implementing a latent profile analysis within a Bayesian framework, informative priors can be used to relax these restrictive assumptions. The present study used a Bayesian LPA of a

community cohort of individuals with PD to determine (a) the presence and (b) the symptom profiles of motor subtypes.

5.2 Methods

5.2.1 Participants

The previously described 'ParkC' sample of 248 individuals with confirmed idiopathic PD was used in this analysis.

5.2.2 Measures

As with Chapters 1 and 3 of this thesis, Part 3 of the Movement-Disorders Society revision of the Unified Parkinson's Disease Ratings Scale (MDS-UPDRS) was used to assess motor symptom severity (Goetz et al., 2008). The previous chapter used Bayesian Structural Equation Modelling (BSEM) to identify that the MDS-UPDRS follows an 8-factor (or symptom) measurement structure. That is, the items in the measure are influenced by 6 motor symptoms (axial, rest tremor, rigidity, akinesia, kinetic tremor, and lower akinesia) and by the laterality of each symptom (left- and right-sided severity). These 8 'latent' symptoms were used in the subtyping analysis. Levodopa Equivalent Dose (LED) was also entered into the model as a control variable.

To account for the possibility of staging process (i.e., different subtypes present at different disease durations), disease duration was not entered into the subtyping model. Rather, the mean disease duration within each subtype was estimated post-hoc. If a staging hypothesis is accurate, we would expect to see marked differences in disease duration between individuals in each subtype. Conversely, if a staging hypothesis is not accurate, then we would expect similar disease durations regardless of subtype.

5.2.3 Statistical Analysis

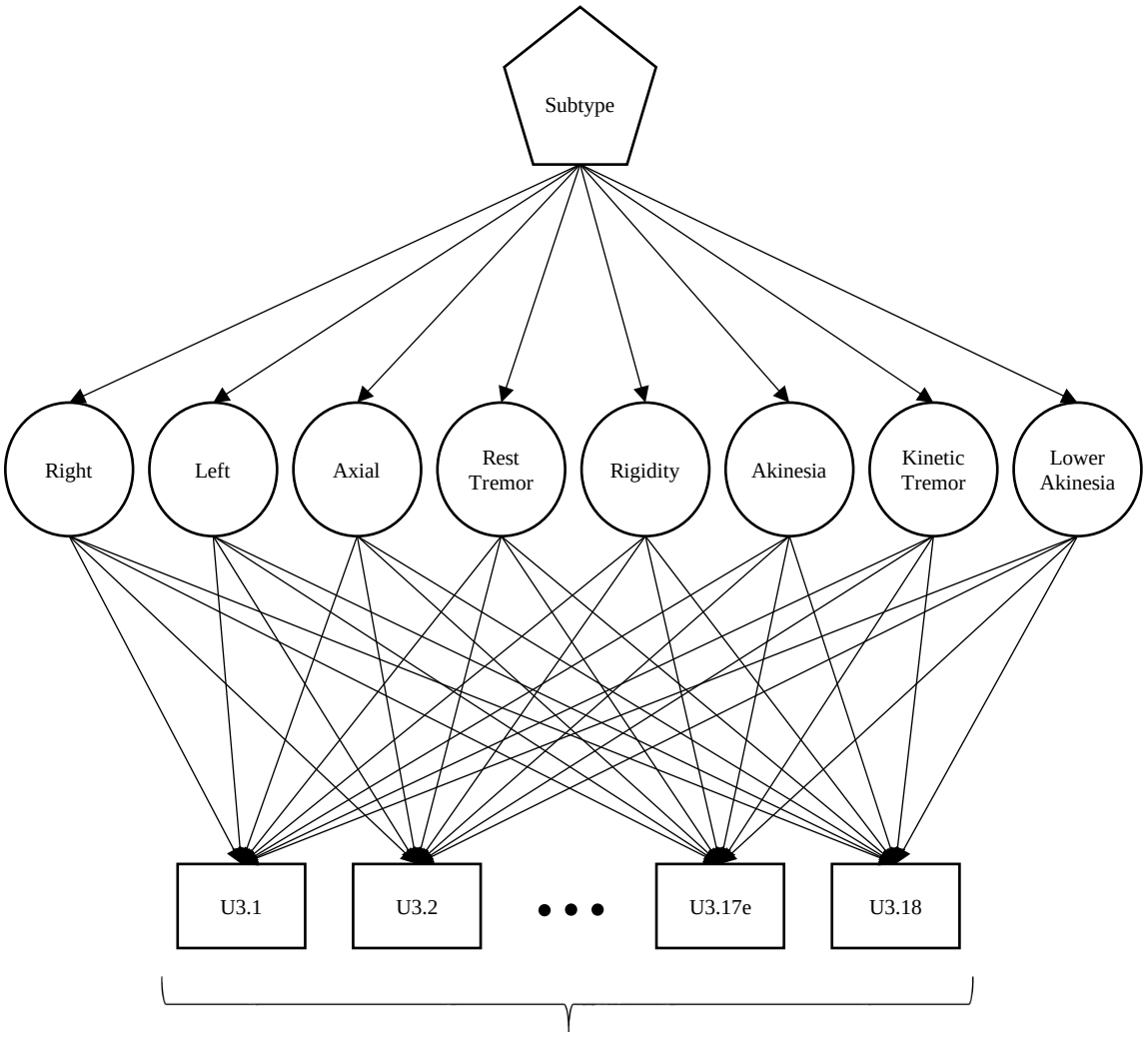
5.2.3.1 Model

The analysis itself can be understood as being comprised of two parts: a measurement model and a subtyping model. The measurement model is the BSEM of the MDS-UPDRS that was developed in the previous chapter. This allows for the latent (error-free) motor symptom variables to be used as outcomes without needing to estimate factor scores in a separate step (which would introduce error in the estimation). This also allows for the effects of the control variables and subtyping process to be evaluated in the same model, circumventing the issues described earlier.

The subtyping model can be understood as the introduction of a categorical latent variable that is indicated by the latent motor symptom factors (diagrammed in Figure 12). Contrast this to the typical use of a *continuous* latent variable (as in traditional factor analysis), where the variations in the continuous level of the latent variable are modelled as being responsible for the variations in the observed severity of its indicators. When a categorical latent variable is estimated, the variation in *category* (i.e., subtype) is used to explain the variation in its indicators (motor symptom factors)

Because the motor symptom factors are unobserved, they do not have a meaningful score to be able to interpret or cluster. Instead, the analysis sets the means of all latent variables in the last class to 0 (also known as the ‘reference class’). The subtypes are then defined by the magnitude of difference in their motor symptom factors from those in the reference (last) class.

Figure 12
Diagram of Subtyping Model



5.2.3.2 Model Priors

All parameters aside from cluster proportions and covariances/correlations were assigned normally distributed priors. Item thresholds and intended factor loadings were given weakly-informative priors of $N(0, 25)$, stating that 99% of values (3 standard deviations) should fall in the range [-15, 15]. Cross-loadings were assigned strongly informative priors of $N(0, 0.01)$, stating that 99% of values should fall in the range [-0.3, 0.3]. The motor symptom factor means within each class (i.e., magnitude of difference from reference class) were given

weakly-informative priors of $N(0, 25)$, stating that 99% of values should fall in the range [-15, 15].

Cluster proportions were given a weakly-informative Dirichlet prior of $D(10_1,\ldots,10_k)$, where k is the number of subtypes in the model (i.e. the 2-Class model had prior $D(10,10)$ whereas the 3-Class model had prior $D(10, 10, 10)$). This prior on cluster proportions is considered weakly-informative, as it is not hypothesising a size for a given cluster but is setting the minimum cluster size that would be considered reasonable (in this case, 10). As, for example, a subtype containing two individuals from a sample of 400 may not be clinically relevant or may simply be an artefact of the sample under study.

Correlations/covariances were given an LKJ prior (named for Lewandowski, Kurowicka, and Joe; the authors who proposed the distribution; Lewandowski et al., 2009) As discussed in Chapter 4, the LKJ prior distribution is controlled by a single parameter η, which determines the strength of prior information about correlations between items/constructs. Both the covariances of the motor symptom factors within each subtype (i.e., local in/dependence) and the residual covariances of observed MDS-UPDRS items were given an informative LKJ prior with $\eta = 20$. This represents strongly informative prior information that the residual covariances and covariances within subtypes (local independence) should be 0.

5.2.3.3 Estimation

All analyses were conducted using RStan. Given the scale and complexity of the model, the analysis was highly computationally intensive (even under a Bayesian paradigm) and required an extended period of sampling for the results to be trustworthy. To facilitate more efficient estimation, extensive work optimising the relevant C++ source code in Stan was undertaken. This is included in Technical Appendices A-E for the interested reader. All

code was reviewed by other core developers of Stan and incorporated into the main program, available from version 2.19 onwards.

Similar to Chapter 3, the technical details behind the estimation of these models is detailed in a series of appendices. The statistical background to constructing a latent profile analysis in a Bayesian model is covered in Appendix G. The difficulties introduced by estimating these models in a Bayesian context are discussed in Appendix H. An annotated version of the Stan syntax used in the present analysis is discussed in Appendix I. The full syntax is presented in Appendix J. The distributions for each parameter in the model are then presented in Appendix K.

Model fit was assessed using the 'Leave One Out' Estimated Log Posterior Density (LOO ELPD; Vehtari et al., 2016), and the Deviance Information Criterion (DIC). Andrew Gelman's approximation of the effective number of parameters (pD; Sturtz, Ligges, & Gelman, 2005) was used to allow the estimation of the DIC in a model with categorical parameters (Lunn, Jackson, Best, Spiegelhalter, & Thomas, 2012). The DIC statistic can be interpreted in a similar fashion to the frequentist Akaike Information Criterion (AIC). These information criteria do not provide evidence about whether a single model fits well but are used to assess whether one model fits the data better than another. Lower values indicate a better fitting model. The DIC can be sensitive to different model parameterisations, and so the LOO ELPD is considered more reliable for model selection (Vehtari et al., 2016).

To determine the number of subtypes, the analysis was run with 1, 2, 3, 4, and 5 class solutions, and the best-fitting model was selected. The best-fitting model was selected based on the previously described model fit statistics and the model entropy. Entropy, in and of itself, is not a fit statistic. Rather, entropy examines a given subtyping analysis and assesses how well it can separate individuals into the respective subtypes. That is, it assesses the level

of confidence that the model has in being able to assign individuals to the ‘correct’ subtype. By considering both the quality of subtyping (entropy) and the fit of the subtyping model (LOO ELPD and DIC), the correct number of subtypes can be selected.

5.3 Results

The fit and entropy statistics for all subtyping models are presented in Table 7. The 1 Class (or “no subtypes”) model demonstrated the poorest fit across all statistics. This provides strong support for the notion of motor subtypes in the sample, as the model hypothesising no subtyping does not describe the observed individuals as well as the models hypothesising subtyping. Given this initial evidence that motor subtypes are present, the next step is to determine the number of subtypes and the motor symptoms that define them.

All subtyping solutions showed more than acceptable levels of entropy (classification certainty), with entropy values greater than the recommended minimum of .80 (Celeux & Soromenho, 1996). This indicates that, regardless of the number of motor subtypes selected, the model can clearly separate individuals into these groups. Given that all subtyping models show good capability to separate individuals, the next step is to assess which number of subtypes (i.e., which manner of separation) best represents the observed patterns of motor symptoms.

Across both model fit statistics (DIC, and LOO ELPD), the 3 Class subtyping model showed the best fit to the observed data. This provides strong evidence of three motor symptom groupings in the sample. The first subtype comprised 26.02% of the sample, and was defined by increased axial, rigidity, akinesia, and kinetic tremor symptoms. The second subtype comprised 29.51% of the sample and was defined by increased right and left lateral symptoms (i.e. ‘overall’ severity) and increased resting tremor symptoms. The third subtype comprised 44.47% of the sample, and shared aspects of both other subtypes.

Table 7
Subtyping Fit Statistics & Entropy

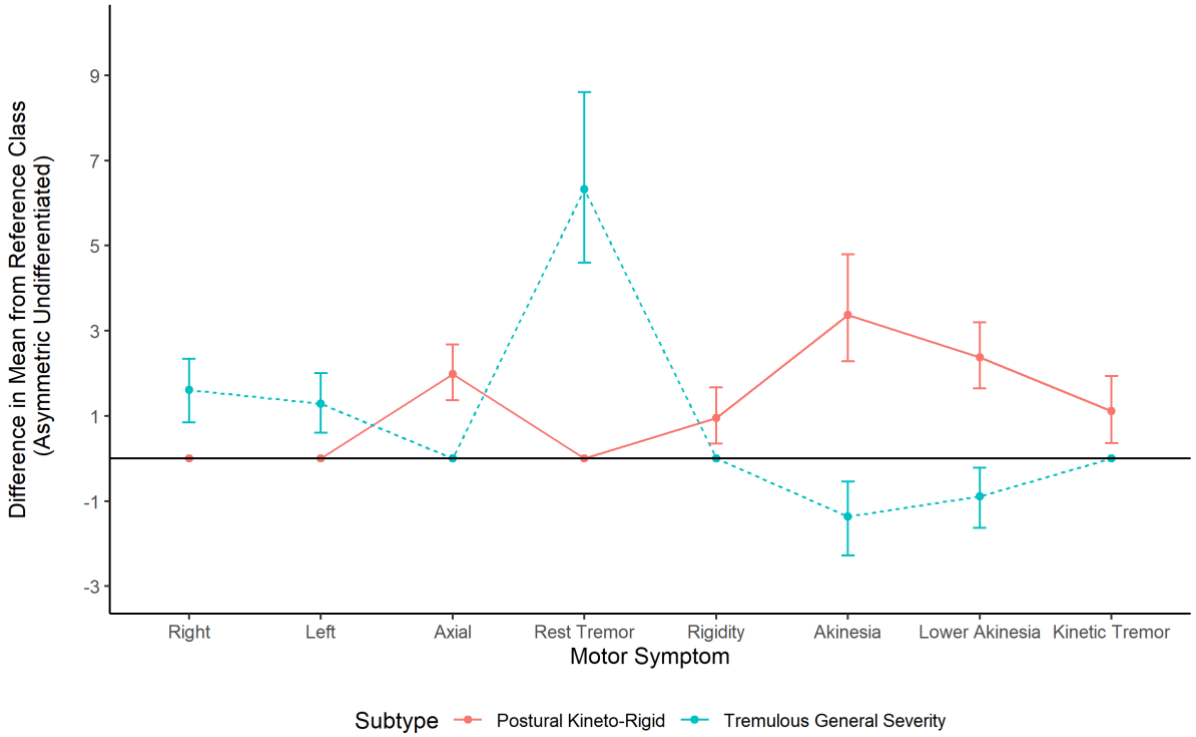
Model	Entropy	DIC	ELPD (LOO)	ELPD Difference [95%CI]
1 Class	-	14926.35	-5665.76	
2 Class	.935	13943.06	-5666.27	-0.51 [-23.59, 22.58]
3 Class	.980	13609.88	-5620.16	46.11 [14.43, 77.78]
4 Class	.966	13718.41	-5625.74	-5.58 [-25.26, 14.10]
5 Class	.965	13759.24	-5622.05	3.683 [-8.66, 16.02]

Note. DIC = Deviance Information Criterion; ELPD (LOO) = Estimated Log-Posterior Density (Leave-One-Out)

Based on these symptom profiles, the first subtype was denoted “Postural Kineto-Rigid” the second denoted “Tremulous General Severity”, and the third “Asymmetric Undifferentiated”. Figure 13 contains plots of these symptom profiles, where the difference in symptom means from the reference class is plotted. For clearer interpretation, only those motor symptoms whose credibility did not contain 0 were plotted (i.e., subtype motor symptom mean differences which were 95% likely not to be 0).

Within each of the three identified subtypes there were also a number of unique correlations between the motor symptoms, providing strong support for the use of Bayesian methodology that allowed these. In the Postural Kineto-Rigid subtype there were positive correlations between axial severity and the right and left severity scores. This positive correlation indicates that as the overall severity of the disease progresses, the magnitude of difference in axial severity between the other subtypes becomes more pronounced. A second positive correlation between the axial and akinetic symptoms indicates a similar pattern, as axial symptoms become more severe so too do akinetic symptoms.

Figure 13
Estimated Motor Symptom Mean Differences of 3 Subtype Solution



This pattern of correlations within the Postural Kineto-Rigid subtype is an interesting contrast in comparison to the Tremulous General Severity type, where no correlations were present at all (see Table 8). A lack of inter-correlations indicates that these individuals differ from the other subtypes only by the severity of their motor symptoms, not the ways in which they are related.

Surprisingly, the correlations in the Asymmetric Undifferentiated subtype indicate the presence of asymmetrical symptom severity. A negative correlation between the left and right severity scores indicates that, as the severity of one increased, the other decreased. This is in clear contrast to the positive correlation seen in the Postural Kineto-Rigid type which indicated that both lateralities varied in severity in the same way. Further, the Asymmetric Undifferentiated individuals also presented with a positive correlation between the upper and

lower akinesia motor symptoms (where no such relationship was present in the Kineto-Rigid type).

Table 8
Motor Symptom Correlations within Symptom Subtypes

Motor Subtype	Motor Symptom Factor							
PKR	1.	2.	3.	4.	5.	6.	7.	8.
1. Right	-							
2. Left	.18	-						
3. Axial	**.29***	**.23***	-					
4. Rest Tremor	-.01	-.02	-.02	-				
5. Rigidity	-.02	.14	.19	-.03	-			
6. Akinesia	.07	.12	**.30***	0	.09	-		
7. Lower Akinesia	.09	.06	.20	-.01	.06	.16	-	
8. Kinetic Tremor	.08	.01	-.01	0	-.07	.09	-.06	-
TGS	1.	2.	3.	4.	5.	6.	7.	8.
1. Right	-							
2. Left	.08	-						
3. Axial	.15	.07	-					
4. Rest Tremor	-.08	-.02	-.02	-				
5. Rigidity	.05	.01	.17	-.01	-			
6. Akinesia	0	0	-.02	-.03	-.04	-		
7. Lower Akinesia	-.03	.01	.11	-.03	-.02	.09	-	
8. Kinetic Tremor	.16	.02	.09	.06	.08	-.08	-.06	-
AU	1.	2.	3.	4.	5.	6.	7.	8.
1. Right	-							
2. Left	**-.22***	-						
3. Axial	.20	.13	-					
4. Rest Tremor	-.03	-.05	-.09	-				
5. Rigidity	-.11	-.01	.18	-.17	-			
6. Akinesia	.02	.02	**.30***	-.08	.11	-		
7. Lower Akinesia	.08	.09	.21	.08	.07	**.28***	-	
8. Kinetic Tremor	-.06	.12	0	.13	-.02	-.06	.04	-

Note. PKR = Postural Kineto-Rigid, TGS = Tremulous General Severity, AU = Asymmetric Undifferentiated
** 95% Crediblity Intervals did not contain 0*

The estimated mean disease durations for individuals most likely to belong to each subtype are presented in Table 9. Consistent with the results of Chapter 2 and other staging research (Nutt, 2016), the subtype with increased postural and rigidity severity (PKR) had the highest disease duration. Interestingly, the TGS type had a higher average disease duration than the AU type. In a staging hypothesis, we would expect that individuals would begin with primarily tremor severity, then an approximate equivalence of tremulous and postural and rigidity symptom severity, followed by primarily postural and rigidity severity. However, the credibility intervals for these estimated means also bear consideration. The upper credibility interval for the mean duration in the AU type was only two months below the lower credibility interval of the TGS type. This implies a likelihood that the 'true' difference in disease duration between the two groups could be as little as two months. This could imply both cross-sectional subtypes and a staging process, where there is a common progression to the PKR type but individual variation in whether the 'beginning' type is TGS or AU. However, when inspecting the same differences in credibility intervals between the TGS and PKR types, the difference is only 23 months (less than two years). A difference this small could simply be an artefact of diagnosis, where individuals with PKR symptoms are being diagnosed earlier than individuals with TGS symptoms.

Table 9
Disease Duration Differences Between PD Motor Subtypes

Subtype	Disease Duration (months) [95%CI]
Postural Kineto-Rigid	109.44 [94.22, 127.16]
Tremulous General Severity	69.10 [65.95, 71.70]
Asymmetric Undifferentiated	59.82 [56.04, 63.73]

A large proportion of individuals originally classified as tremor dominant were re-classified as asymmetric-undifferentiated under the new approach (see Table 10).

Table 10
Differences in Subtype Classification Between Stebbins et al. (2013) and the Proposed Approach

	PKR	AU	TGS
PIGD	60	21	21
IND	6	18	3
TD	6	91	22

Note. PIGD = Postural Instability and Gait Difficulty; IND = Indeterminate; TD = Tremor Dominant; PKR = Postural Kineto-Rigid; AU = Asymmetric-Undifferentiated; TGS = Tremulous General Severity

The reasons for this can be explained by inspecting the MDS-UPDRS item means within this group of individuals (see Table 11). While these individuals had low levels of severity on the items used to assess to classify individuals as PIGD (3.10, 3.11, and 3.12), they had higher levels of severity on the other, non-postural, MDS-UPDRS items. In effect, while the tremor severity of these individuals was greater than their postural severity (resulting in their TD classification), it was not greater than their severity of rigid and/or akinetic symptoms. In this way, the proposed approach to subtyping allows for a more comprehensive accounting of individual variations in motor severity, by capturing the differences in all motor symptoms, not just tremor and postural.

Table 11
MDS-UPDRS Item Means for TD Individuals Re-Classified as AU

MDS-UPDRS Part 3 Item	TD -> AU (N = 91)
3.1	0.62
3.2	1.22
3.3a	1.2
3.3b	1.22
3.3c	1
3.3d	0.88
3.3e	0.87
3.4a	1.38
3.4b	1.45
3.5a	1.09
3.5b	1.17
3.6a	1.07
3.6b	1.32
3.7a	1.11
3.7b	1.17
3.8a	0.78
3.8b	0.87
3.9	0.21
3.10	0.77
3.11	0.07
3.12	0.38
3.13	0.5
3.14	1.08
3.15a	0.8
3.15b	0.81
3.16a	0.74
3.16b	0.94
3.17a	1.18
3.17b	1.05
3.17c	0.22
3.17d	0.28
3.17e	0.11
3.18	2.43

5.4 Discussion

After controlling for individual differences in antiparkinsonian medication, as well as removing motor symptom measurement error, the analyses indicated that there were three PD motor symptom subtypes in the present sample. The first group of individuals, termed "Postural Kineto-Rigid" presented with increased severity of akinesia, rigidity, axial/postural, and kinetic tremor symptoms. The second group, termed "Tremulous General Severity" presented with increased overall severity and markedly increased resting tremor severity. The third group was undifferentiated, showing features of both groups. A further interesting result that emerged was the clear distinction in relationships with resting and postural/kinetic tremors. Where resting tremors were clearly indicative of their own subtype, postural/kinetic tremors appeared more closely associated with the Postural Kineto-Rigid subtype.

Individually, these are all results that have been seen in previous, separate, studies. Akinesia, rigidity, and axial/postural symptoms have all previously been (separately) proposed as a means of grouping individuals with PD. These symptoms, however, have never been applied in combination while assessing subgroupings in PD. This may provide some explanation for the present results differing to those of previous explorations. Previous studies have all separately identified a subgroup with akinesia, rigidity, or axial/postural symptoms as being dominant (i.e., of greater severity than the others). It may have been that these studies were unknowingly capturing a group with increased severity of all three. This single subtype is consistent with the underlying neuropathology shared by these symptoms. Postural, rigidity, and akinesia motor symptoms have all demonstrated links to the mesencephalic locomotor region, specifically the pedunculopontine nucleus (PPN; Pahapill & Lozano, 2000).

The PPN is a neuronal structure in the upper brainstem which receives projections from basal ganglia nuclei and in turn projects to the spinal cord (Tubert, Galtieri, & Surmeier,

2018). This interconnection allows the PPN to act as a 'relay station' of sorts, modulating the synaptic activity (and resulting movements) between the two structures (Tubert et al., 2018). The PPN has been strongly implicated in the degeneration of postural control and balance in PD (Ferraye et al., 2010), with primate lesion models further demonstrating the inducement of axial, rigidity, and hypokinetic symptoms with no corresponding onset of tremor (N. Jenkinson et al., 2009). This pattern of increased severity in axial, rigidity, and akinetic symptoms with no difference in tremor symptoms bears a striking similarity to the Postural Kineto-Rigid type identified in the present study. Taken together, this would suggest that the Postural Kineto-Rigid subtype is identifying individuals with primarily (or more extensive) degeneration of the PPN. This common neuropathology could explain why previous subtyping studies have identified various combinations of postural, akinetic, and rigid subtypes – each were capturing an incomplete conceptualisation of the Postural Kineto-Rigid type.

Previous research has also distinguished between resting and postural/kinetic tremors and their relationships with (other) PD subtypes – identifying that postural tremors were more severe in an akinetic-rigid than tremor-dominant subtype (Deuschl et al., 2000; Helmich, Janssen, Oyen, Bloem, & Toni, 2011). However, these authors did not postulate whether these kinetic tremors could be a defining feature of a subtype. In previous PD subtyping studies, kinetic tremors have also been combined with resting tremors as indicators of an overall "tremor" construct (Lewis et al., 2005; Reijnders et al., 2009). However, the results of the present study indicate that this could be misrepresenting the ways in which the symptoms present in PD. These findings support the distinction between the two types of tremor symptoms, and the neurological mechanisms involved. Given that kinetic tremors were not seen as an aspect of the "Tremor" type, seeming more likely to belong to the "postural

kineto-rigid" type, this suggests that the two tremor symptoms could be indicative of different neuropathologies.

Unfortunately, the assessment of whether disease duration differences between the subtypes could indicate the presence of a staging process of subtyping was inconclusive. While there were differences in disease duration, the credibility intervals indicated that the true differences could be negligible. This is further complicated by the reliance on time since diagnosis as an indicator of disease duration. The time taken to diagnose an individual with PD, and rule out misdiagnoses, can be influenced by the particular symptoms with which an individual presents (Tolosa et al., 2006). As such, time since diagnosis is not always a reliable or accurate indicator of the length of time that an individual has had PD. Overall, it is difficult to make confident inferences about the likelihood of a staging hypothesis without a more accurate measure of disease duration or longitudinal research exploring membership of the newly-derived subtypes over time.

Future research should explore this likelihood of a staging hypothesis. The recommended statistical approach for this kind of hypothesis is a longitudinal extension of latent profile analysis called 'latent transition analysis' (LTA; Lanza, Bray, & Collins, 2013). Where latent profile analysis assesses the presence of subtypes within a cross-sectional group, LTA extends this to explore the probability of an individual transitioning between these subtypes across subsequent assessments (Lanza et al., 2013). By using an LTA-based approach, studies could explicitly test a subtype staging hypothesis by assessing whether there is a common 'direction' of subtyping membership change (i.e. from a tremor type to a postural/akinesia/rigidity type).

However, these analyses come with a greater sample size requirement than most longitudinal analyses (such as the growth curve model of Chapter 2). This is because LTA

requires a sufficient sample size to perform the subtyping analysis at each assessment time (Oberski, 2016). While this does represent somewhat of a ‘design burden’, the results could provide great insight into the likelihood of subtyping staging in PD.

The present results can, effectively, be considered a ‘tying together’ of previously disparate findings. A likely reason for this inconsistency in past results is the inconsistency in the methods that have been used. As described throughout the present thesis, the approaches to PD motor symptom measurement and subtyping have varied greatly throughout the literature. In addition to this, the assumptions made by these methods (i.e., local independence) have also not generally been the most substantively appropriate. By taking a more comprehensive and methodologically robust approach to the subtyping process the present study has been able to identify subtypes that have only been partially captured in previous research. However, for these subtypes to have clinical utility beyond the theoretical, it is important to identify how quality of life and disease experience differs for individuals within these groups.

5.5 Chapter 5 – Summary

This chapter derived and presented a Bayesian extension of LPA for motor subtyping in PD. The proposed method allows for the analysis to account for relationships between motor symptoms within each subtype. The proposed method can also account for disease duration, antiparkinsonian medication, and measurement error while subtyping.

Three subtypes were identified. The first subtype was denoted “Postural Kineto-Rigid”, where individuals presented with increased severity of postural, akinetic, and rigidity symptoms. The second subtype was denoted “Tremulous General Severity”, where individuals presented with increased overall severity and markedly increased severity of resting tremor symptoms. The third subtype was denoted “Asymmetric Undifferentiated”.

Individuals in this subtype presented with a combination of the motor symptoms of the other two subtypes, but with a negative correlation between the severity of symptoms on either laterality of the body. That is, as the severity of symptoms on one side of the body increased, the severity on the other side decreased.

The comorbid severity of the symptoms in the Postural Kineto-Rigid subtype would suggest that this subtype is associated with increased degeneration of the pedunculopontine nucleus (PPN). This common neuropathology would explain why previous research has separately identified subtypes that are indicated by increased postural or akinetic-rigid severity – as both subtypes were poorly capturing the presence of the Postural Kineto-Rigid subtype.

Now that PD subtypes have been identified in a methodologically rigorous and substantively appropriate manner, it is important to explore their clinical relevance.

6 Chapter 6 – Quality of Life Experiences in PD Subtypes

6.1 Introduction

Identifying the predictors of quality of life (QoL) differences in people with Parkinson's disease (PD) is of continuing interest in the wider literature (van Uem et al., 2016). Given the (currently) incurable nature of the disease, there is a need to understand and mitigate disease impacts on an individual's daily functioning and quality of life. As PD can present with a broad range of cognitive, autonomic, and physical impairments, individuals experience markedly lower quality of life than the general population (Schrag, Jahanshahi, & Quinn, 2000). While physical symptoms can greatly reduce the ability to carry out daily tasks, often the reduction in social and interpersonal functioning can be more profoundly debilitating (Dauwerse, Hendrikx, Schipper, Struiksma, & Abma, 2014). For example, individuals with PD have identified that a reduction in facial expressiveness and verbal fluency has reduced their participation in conversation and led to greater feelings of isolation (Dauwerse et al., 2014). In addition, the loss of autonomy and increasing dependence on caregivers or spouses as the disease progresses in severity can give rise to feelings of shame or embarrassment (Caap-Ahlgren & Lannerheim, 2002). Being able to identify which individuals are likely to experience poorer quality of life would allow for the implementation of targeted interventions to improve this reduction in QoL.

A systematic review by Soh, Morris, and McGinley (2011) of the predictors of QoL in PD identified that gait impairments were the most consistently related motor symptom (Soh et al., 2011). This pattern of results is also consistent with previous research exploring the relationships between PD motor subtypes and QoL. Cross-sectional studies conducted by Hariz and Forsgren (2011) and G. W. Duncan et al. (2014) showed significantly poorer QoL in individuals belonging to the postural instability and gait difficulty (PIGD) subtype, compared to those individuals in the tremor dominant (TD) or mixed type. Similarly, both

Schrag et al. (2000) and Andreadou et al. (2011) identified significantly poorer QoL in the akinetic-rigid (AR) type compared to the TD or mixed types. This similarity of relationships between the PIGD and AR subtypes is consistent with the results of the previous chapter, which suggested that the two subtypes are both representing the same underlying pathology (as part of the postural kineto-rigid type). As such, we would expect to see poorer QoL in the PKR type.

As discussed in the previous chapter, one of the benefits of latent profile analysis (LPA) is that it estimates each individual's probability of belonging to each subtype. These individual probabilities can be used in subsequent analyses, such as the present study (Neely-Barnes, 2010). As was the goal in the previous chapter, this has the benefit of improving the clinical relevance of the present study. Rarely in practice will it be possible to determine with certainty the motor subtype to which an individual belongs. By using the estimated probabilities from an LPA, the analysis is then assessing how these QoL outcomes differ as individuals are more or less likely to belong to a given subtype (as compared to the reference Asymmetric Undifferentiated subtype). For example, as an individual becomes more likely to belong to the PKR type (as their postural, akinetic, and rigidity symptoms become more severe than their other motor symptoms), how do their QoL outcomes change?

The present study assessed the relationship between the probability of belonging to the PKR subtype or the tremulous general severity (TG) subtype and QoL in PD. It was hypothesised that as an individual's probability of belonging to the PKR type increases, their QoL worsens. It was further hypothesised that this relationship is specific to the PKR type, and that there will be no relationship between the likelihood of belonging to the TS type and QoL in PD.

6.2 Methods

6.2.1 Participants

The previously described ‘ParkC’ sample of 248 individuals with confirmed idiopathic PD were included in this analysis.

6.2.2 Measures

The 39-item Parkinson’s Disease Questionnaire (PDQ-39) was used as the measure of QoL (Peto, Jenkinson, & Fitzpatrick, 1998). The PDQ-39 is a disease-specific measure of QoL, assessing the experiences and impact of health problems specific to PD (e.g., hallucinations, postural instability, and involuntary drooling). The PDQ-39 is the most commonly-used and psychometrically evaluated measure of QoL in PD, with well-established reliability and validity (Martinez-Martin et al., 2011; Soh et al., 2011). The PDQ-39 has eight subscales assessing different dimensions of symptom experience in PD: mobility, activities of daily living, emotional well-being, stigma, social support, cognitions, communication, and bodily discomfort (Peto et al., 1998). Higher scores on these subscales indicate greater impairment on the respective domain.

While the PDQ-39 also has a summary score (‘single index’) representing the overall QoL experience in PD (C. Jenkinson, Fitzpatrick, Peto, Greenhall, & Hyman, 1997), the individual domain scores were used in the present analysis to more accurately assess how the experience of PD differs between the motor subtypes derived in the previous chapter.

6.2.3 Statistical Analysis

The hypotheses were tested using path analysis with maximum-likelihood estimation, as implemented in Mplus version 8.2. A single path model was tested, with each of the PDQ-39 domains entered as correlated outcomes. Standard errors and bias-corrected confidence intervals were estimated using 20 000 bootstrap samples and coefficients evaluated at an

alpha level of .05. An interested reader may note that the present analysis is conducted under a traditional (Frequentist) framework, rather than the Bayesian methods used in the previous chapters. As the current analysis does not require adjustments for residual covariances or assumptions of local independence, a Bayesian methodology was not required.

It should be noted that it is redundant to include all three estimated subtype probabilities as predictors. Given that the three probabilities sum to one, as long as two of them are included, the other is implicitly known. This induces a perfect multicollinearity between the predictors, as the value of one probability is wholly dependent on the values of the others. By only entering two of the probabilities, this sets the third group as a 'reference group'. For this model, the Asymmetric-Undifferentiated (AU) group was selected as the reference, so the relationships with the PKR and TGS subtypes could be compared with the relationships with the TD and PIGD subtypes that have been previously seen in the literature.

A single path model was tested. For each individual, their estimated probabilities of belonging to the TS and PKR types were included as predictors. As these probabilities had been estimated from the subtyping analysis in the previous chapter, they represent the probability of belonging to a given subtype (i.e., TS or PKR) rather than the reference class (AU). The eight subscales of the PDQ-39 were included as outcomes. Disease duration (time since diagnosis) and levodopa equivalent dose (LED; Tomlinson et al., 2010) were included as control variables.

6.3 Results

The results of the path analysis indicate partial support for the hypotheses. The probability of belonging to the PKR type was related to some, but not all, of the QoL outcomes. The analysis indicated that for each 10% increase in an individual's probability of belonging to the PKR type, their mobility, activities of daily living, cognition, and communication significantly worsened (see Table 12). There was, however, no relationship between the probability of belonging to the PKR type and experiences of emotional well-being, stigma, bodily discomfort, or social support.

In contrast, there was no relationship between the probability of belonging to the TS type and experiences in any of the QoL domains. This pattern of results indicates that there is clear differentiation between the PKR and TS motor subtypes in how they experience PD, but that this differentiation does not extend to all aspects of the QoL. This pattern of findings is further supported by the R^2 values for each of the outcomes in the model. The present model with motor subtype probabilities as predictors and disease duration and LED as control variables can account for a large, significant, proportion of variance in only some of the QoL domains (see Table 13). That is, individual variation in the experiences of QoL can be explained by the motor subtype to which an individual is likely to belong, but only for select domains of QoL.

Table 12
Path Coefficients with Bootstrapped Bias-Corrected Confidence Intervals from Path Model with Correlated Outcomes

	Estimate	[95%CI]	*p*-value
Mobility			
TS	0.50	[-0.13, 1.18]	.134
PKR	1.59	[0.66, 2.59]	< .001
Activities of Daily Living			
TS	0.14	[-0.36, 0.70]	.595
PKR	1.88	[1.12, 2.72]	< .001
Emotional Well-being			
TS	-0.13	[-0.61, 0.36]	.600
PKR	0.10	[-0.50, 0.74]	.757
Stigma			
TS	-0.20	[-0.70, 0.31]	.433
PKR	-0.44	[-1.09, 0.26]	.202
Cognitions			
TS	-0.07	[-0.57, 0.45]	.780
PKR	0.71	[0.09, 1.39]	.031
Communication			
TS	0.14	[-0.35, 0.68]	.584
PKR	1.33	[0.66, 2.06]	< .001
Bodily Discomfort			
TS	-0.33	[-1.01, 0.33]	.330
PKR	-0.28	[-1.10, 0.53]	.502
Social Support			
TS	0.07	[-0.44, 0.60]	.796
PKR	0.11	[-0.44, 0.67]	.706

Table 13
R^2 Values and Significance Levels for PDQ-39 Domains in Path Model with Correlated Outcomes

PDQ-39 Domain	R^2	*p*-value
Mobility	.16	< .001
Activities of Daily Living	.21	< .001
Emotional Well-being	0	.764
Stigma	.03	.194
Cognitions	.08	.040
Communication	.18	< .001
Bodily Discomfort	.02	.255
Social Support	.03	.214

6.4 Discussion

The present study aimed to assess the differences in QoL between the newly-derived motor subtypes. The study hypothesised that the postural kineto-rigid (PKR) type would experience poorer QoL across all domains, while the tremulous general severity (TGS) type would show no significant associations. These hypotheses, however, were only partially supported. While the PKR type did show significantly reduced QoL, this was only on four (of eight) domains: mobility, activities of daily living, cognition, and communication. In support of the hypotheses, the TGS type did not show any associations with the various QoL domains. Taken together, this indicates that there are clinical experiences of PD unique to the PKR motor subtype.

Of further interest is why only select QoL domains were reduced within the PKR type. Inspecting the PDQ-39 items within each of these domains would suggest that it is the

severity of the postural, akinesia, and rigidity symptoms themselves that are responsible. The Mobility subscale contains items assessing difficulties walking long distances, carrying bags, and a fear of falls (C. Jenkinson et al., 1997). Given that the PKR type is indicated by motor symptoms affecting balance and ease of movement (i.e., postural stability and joint rigidity), it is unsurprising that individuals with increased severity of these symptoms would have reduced levels of mobility. A similar relationship holds for the Activities of Daily Living subscale, which assesses activities such as dressing, writing, and preparing food (C. Jenkinson et al., 1997). These are all activities that involve some form of fine motor control. Taylor Tavares et al. (2005) showed a strong, significant, relationship between bradykinesia severity and performance on fine motor control tasks. Given that the PKR type is partly characterised by bradykinetic severity, this could account for the reduction in ability to carry out the activities of daily living that are being assessed. This would also seem to hold true for the Communication subscale. The items in this scale assess difficulties with speech and being understood (C. Jenkinson et al., 1997). Both facial expression and speech production can be greatly reduced as a function of akinetic/bradykinetic severity in PD (Cantiniaux et al., 2010). With higher levels of these motor symptoms, it is likely individuals may have difficulties communicating effectively and interacting with others.

The Cognition scale, however, does not follow this pattern of motor symptom association. The Cognition scale of the PDQ-39 assesses symptoms such as difficulties sleeping, concentrating, and the experience of hallucinations (C. Jenkinson et al., 1997). While these symptoms are not associated with the motor symptoms comprising the PKR type, they have previously been identified as impaired in the PIGD subtype (G. W. Duncan et al., 2014). G. W. Duncan et al. (2014) attributed the severity of these symptoms to the increased degeneration of the cholinergic system commonly attributed to the PIGD subtype. As the relationship with the PDQ-39 Cognition scale was seen in both the PIGD type and in the

newly-derived PKR type, this could further suggest that the PIGD type is simply a subset of the PKR motor subtype.

The presence of significant QoL relationships only with the PKR subtype also has implications for the possibility of a staging process for these subtypes. Given that the PKR subtype was shown in the previous chapter to have the highest average disease duration, it is also consistent that the QoL of these individuals would be the lowest (as their disease is the most advanced in its progression).

There are two primary limitations of this study that should be considered. Firstly, given the broad nature of the construct under examination (QoL) it is difficult to examine any single domain in depth without extensive post-hoc comparisons. The present study is, at best, able to provide a general indication of the salient differences in disease experience between the derived motor subtypes. Secondly, the use of a cross-sectional methodology limits the depth of inference available. Without a longitudinal study, it is difficult to say with certainty how these QoL symptoms will change with disease progression, as well how this may differ between the motor subtypes.

Overall, the present study provides initial evidence for clinical differences in QoL and disease experience between the PKR and TGS motor subtypes. Being able to identify individuals at risk of poorer QoL outcomes using a simple motor examination (and subsequent subtyping) could provide a clinically useful tool for targeting interventions. Having a clear, clinical, application for motor subtyping would allow it to move from the research domain to a tool that can feasibly be used to improve patient outcomes. If, for example, an individual was identified to be likely to belong to the PKR type, a clinician would be able to direct that individual towards additional interventions (such as yoga or tai chi) as well as additional social support networks (such as support groups).

6.5 Chapter 6 – Summary

This chapter explored the differences in disease experience (using quality of life as a proxy) between the newly-derived PD motor subtypes. The analysis identified that there were no significant changes in QoL as the probability of belonging to the Tremulous General Severity (TGS) increased. In contrast, as the probability of belonging to the Postural Kineto-Rigid (PKR) type increased, quality of life worsened in four domains: mobility, activities of daily living, cognition, and communication.

These results provide some validation for the proposed idea that the PIGD and akinetic-rigid subtypes are both facets of the PKR type, as both have demonstrated similar relationships in the past. Of greater importance, these results demonstrate a clear clinical relevance for these newly-derived motor subtypes. Being able to identify individuals at risk of poorer health outcomes provides an opportunity to implement targeted preventative measures. The ability to offer improved patient care may be what is needed for the field of PD subtyping to progress from a research topic to a clinical practice.

7 Chapter 7 - General Conclusion

7.1 Summary of Findings

7.1.1 Study 1 (Chapter 2)

Study 1 applied latent growth modelling to measurements of 236 individuals taken over six years. This study sought to characterise the way that tremor and postural symptoms change in severity over time in idiopathic PD.

The analyses identified that, after accounting for changes in antiparkinsonian medication, tremor symptoms showed no significant yearly change in severity. Further, this (negligible) amount of change in severity did not significantly vary between individuals. These results indicate that, on average, tremor symptoms showed no significant changes in severity over time, and that this was consistent throughout the sample. It should, however, be noted that the amount of change in tremor severity is only non-significant after accounting for individual differences in antiparkinsonian medication. This implies that some individuals may only appear to have higher levels of tremor severity because they are on a different medication dosage to other individuals.

In contrast to the tremor symptom findings, the analyses identified that postural symptoms showed a significant yearly increase in severity. The results further indicated that this amount of change in severity significantly varied between individuals. There was no relationship between the changes in antiparkinsonian medication and the changes in postural symptom severity. These findings imply that while, on average, postural symptom severity increases in severity over time, this is not related to antiparkinsonian medication. These results further imply that there is marked individual variability in these changes over time, such that some individuals change in severity at a more rapid pace than others.

Further analysis of the changes in tremor and postural severity identified that the initial severity and rate of change in tremor symptoms was not related to the initial severity or rate of change in postural symptoms. In effect, the progression of tremor severity was not related to the progression of postural severity in these individuals with PD.

Based on these findings, Study 1 concluded that Jankovic et al.'s (1990) TD and PIGD classifications were not truly reflective of PD subtypes, as the magnitude of difference between tremor and postural symptoms was shown to be a function of disease duration. Study 1 further concluded that investigation into a staging hypothesis of PD subtypes was needed.

7.1.2 Study 2 (Chapter 4)

Study 2 applied Bayesian structural equation modelling to motor assessments of 248 individuals with idiopathic PD. Study 2 aimed to develop a model of Parkinsonian motor symptom measurement that was more reflective of the clinical presentation of the disease than that of previous models.

Study 2 identified six primary motor symptoms being assessed by the motor examination of the MDS-UPDRS: axial, resting tremor, postural/kinetic tremor, rigidity, akinesia, and lower-body akinesia. The analyses further identified that the severity of these six motor symptoms also differed by laterality. That is, the severity of a given motor symptom was not always consistent across both sides of an individual's body. This model of measurement was invariant in a separate sample of 423 de novo individuals with PD. Finding that the model was invariant across multiple samples, and invariant in an unmedicated sample, provides a strong level of support.

Of particular importance was the finding that the rating on a given MDS-UPDRS item can be influenced by multiple motor symptoms. For example, the rating on MDS-UPDRS item 3.10 (gait) may not only represent the severity of postural symptoms; but also, rigidity

and akinesia in the lower limbs. In effect, while each MDS-UPDRS item may be designed to assess the severity of a particular motor symptom (e.g., rigidity), this assessment is being influenced by the severity of other motor symptoms (e.g., akinesia). Consequently, given that individual items are imperfect measures of a motor symptom, the sum or mean of these items will also be an imperfect measure. This has significant implications for the assessment of PD motor symptoms in both clinical and research settings.

Based on these findings, Study 2 concluded that the previous subtyping analyses using sum or mean scores to represent motor symptom severity could be biased by the inclusion of these poor measurements.

7.1.3 Study 3 (Chapter 5)

Study 3 proposed a Bayesian extension of LPA to explore the presence of motor subtypes of PD in a sample of 248 individuals. Study 3 sought to develop and propose a method of PD subtyping that better represents the clinical presentations of PD heterogeneity as well as being more statistically powerful than current methods. Of clinical relevance, the method is the first of its kind to allow for the measured symptoms within each subtype to be related. Of statistical relevance, the proposed method is the first that can simultaneously account for measurement error, antiparkinsonian medication, missing data, and correctly handle categorical data.

The analyses identified that three motor subtypes were present. The first subtype was denoted “Postural Kineto-Rigid” (PKR) and was defined by increased severity of postural, akinetic/bradykinetic, and rigidity symptoms. The second subtype was denoted “Tremulous General Severity” (TGS) and was defined by increased overall severity and markedly increased resting tremor severity. The third subtype was denoted “Asymmetric Undifferentiated” (AU) and was defined by a mixture of the symptoms of the first two

subtypes, but with a negative correlation between the severity of symptoms of either laterality. In other words, an increased level of severity on one side of the body was associated with a decreased level of severity on the other. Assessing the differences in mean disease duration between each of the estimated subtypes did not provide conclusive evidence either for or against a staging process of PD subtypes.

The methods proposed by Study 3 demonstrated clear benefits over those of previous studies. The presence of correlations between motor symptoms within each subtype provided strong support for this conclusion. Consequently, the results of Study 3 suggest that previous studies assuming symptoms would be unrelated within each subtype would likely have arrived at biased conclusions.

7.1.4 Study 4 (Chapter 6)

Study 4 used path analysis to demonstrate a clinically relevant difference in quality of life between the motor subtypes that were derived in the previous chapter. For each individual, their estimated probabilities of belonging to the TGS or PKR type were saved from the subtyping analysis in Study 3 and used as predictors. Using these probabilities as predictors instead of the final subtype membership allowed for more precise estimates of the relationship between the subtypes and differences in quality of life.

The analyses identified that the individuals belonging to the TGS and AU motor subtypes did not significantly differ in their experiences of quality of life. Individuals belonging to the PKR type, however, experienced significantly poorer quality of life across several domains than individuals in the AU type. The largest differences were seen in activities of daily living, defined by the ability to complete tasks such as dressing and preparing meals. Similar differences were seen in mobility, defined by the ability to complete tasks such as walking and carrying bags. Next, were differences in communication, which

assesses the ability to speak and communicate. The smallest differences were seen in cognition, which assesses symptoms such as difficulties sleeping and concentrating.

These significant differences, and their presence only in the PKR type, indicates a clear clinical application for PD motor subtyping in accordance with this thesis. The ability to identify individuals at greater risk of poor quality of life (and the associated negative outcomes) through a simple motor examination could be of clear prognostic value in clinical practice. Thus, Study 4 concluded that the present results indicate a proof-of-concept for the potential of motor subtyping as a tool for clinical practice.

7.2 Implications of the Present Thesis

One of the key conclusions drawn by this thesis is that current methods of PD subtyping are strongly limited. These limitations are present in both forms of PD subtyping that are currently in use: *a priori* and data-driven methods. The present thesis demonstrated that these limitations were present in three primary areas: differential rates of change in severity, limitations in measurement, and limitations in statistical methods.

7.2.1 Differential Rates of Change in Symptom Severity

Jankovic et al.'s (1990) TD and PIGD *a priori* classifications were highly influential at their introduction and their use continues to the present day. Jankovic et al.'s (1990) classifications are based on the comparison of tremor and postural symptom severity. However, Study 1 demonstrated that tremor and postural symptoms change in severity at different rates. This difference in the rate of change in severity is such that, given enough time, most individuals will exhibit more severe postural symptoms than tremor symptoms. In effect, this implies that TD or PIGD membership is more indicative of disease duration than disease subtypes. Being able to quantify disease progression through these groupings would

have clear clinical utility, but this is not how the TD/PIGD groupings are used. This finding has strong implications for the PD subtyping field, as well as PD research as a whole.

This concern is not unique to those studies using the TD and PIGD classifications. Given that tremor and postural symptoms change in severity at different rates, this may also be the case for other Parkinsonian symptoms. As such, other forms of *a priori* subtyping classification may be affected. A prime example of this would be the akinetic-rigid (AR) subtyping proposed by Schiess et al. (2000) and later refined by Lewis et al. (2005). While not as popular as the TD and PIGD classifications, the AR subtype still sees use in current research (Guan et al., 2017; Zhang et al., 2015). The use of this classification system, and the results of studies that have previously applied this system, should also be re-evaluated in longitudinal context.

The proper handling of differential rates of change in severity is also a concern for data-driven approaches to classification. Currently, data-driven subtyping studies, such as that conducted by van Rooden et al. (2011), attempt to control for differences in disease duration separate to the subtyping analysis. van Rooden et al. (2011) approached this task by saving the residuals from a linear regression for each subtyping variable with disease duration as the outcome. The estimated effects of disease duration in these analyses will be inaccurate for two reasons. Firstly, by estimating separate regressions for each outcome, van Rooden et al. (2011) are unable to account for the correlations between the outcomes. If the outcomes are correlated, and the regression is unable to account for this, the estimated relationships could be inflated. Secondly, by assessing the effects of disease duration separately to the effects of disease subtype, van Rooden et al. (2011) are not able to accurately distinguish the two. Without accounting for disease duration in the same analysis as the subtyping process, the effects of each on differences in symptom severity could be over- or under-estimated.

Overall, the finding that tremor and postural symptoms change in severity at different rates highlights the poor control of disease duration in subtyping studies to date. This finding casts doubt on the validity and accuracy of the present state of the subtyping field.

7.2.2 Limitations in Measurement

A second, significant, finding of the present thesis concerns the ability to measure motor symptom severity. Currently, studies wishing to use the severity of different motor symptoms (e.g., tremor, rigidity) in a subtyping analysis will take the sum or mean of several MDS-UPDRS items (of their choosing). However, the findings of Study 2 indicate that this not appropriate.

There are, of course, reliability and validity concerns inherent to studies using different combinations of MDS-UPDRS items to represent the same motor symptom. Take the subtyping studies conducted by Damholdt, Shevlin, Borghammer, Larsen, and Ostergaard (2012) and Reijnders et al. (2009). In these studies, Reijnders et al. (2009) scored tremor severity using UPDRS items 16, 20, and 21; whereas Damholdt et al. (2012) scored tremor severity using UPDRS items 20 and 21 alone. As these two studies have used different items, their ‘tremor’ constructs are not directly comparable. Of further concern, as these choices of UPDRS items were not based on a validated measurement or factor structure, it cannot be said with confidence that either study is truly measuring tremor severity. The measurement structure presented by Study 2 offers a possible remedy to this.

Study 2 presents a measurement structure of the MDS-UPDRS that has been replicated in two samples from different countries (the ParkC cohort based in Australia and the PPMI cohort based in America). By providing this additional level of rigour in evaluation, the present thesis aims to reduce the barriers to uptake of the model in future PD studies. Given the inconsistent approaches to scoring severity of individual motor symptoms in PD

research, the provision of a reliable model could begin to introduce some much-needed consistency to the field.

Beyond the choice of MDS-UPDRS items, the results of Study 2 highlight a more concerning flaw with current PD motor symptom measurement. The analytic methods of Study 2 (Bayesian structural equation modelling; BSEM) allowed for a depth of evaluation that has not previously been possible. Through the application of BSEM, Study 2 identified that individual MDS-UPDRS items are not able to perfectly isolate and measure the motor symptom for which they were developed. While this may be an intuitive concept, it has not been accounted for when researchers score motor symptom severity. By simply taking the sum or mean of MDS-UPDRS items, studies are assuming that each of those items are perfect measures of the motor symptom of interest. The findings of Study 2 show that this approach is likely to introduce bias, and that the resulting sums or means risk being inaccurate representations of the severity of the motor symptoms under study.

This finding has direct relevance for PD subtyping. Mean/sum scores of MDS-UPDRS items are often the basis upon which individuals are grouped into different subtypes (Damholdt et al., 2012; Reijnders et al., 2009). If these mean/sum scores are not accurate representations of the motor symptoms that they are intended to assess, the subtypes that are derived conditional on these scores will also likely not be accurate. In effect, if part of the individual differences in these sum/mean scores are due to measurement error, than this measurement error could bias a subsequent analysis attempt to subtype individuals base on these individual differences.

This finding, however, also has implications for the wider PD research field. The use of mean/sum scores of MDS-UPDRS items is not unique to the domain of PD subtyping. These mean/sum scores have been used as predictors and/or outcomes in a range of studies,

including randomised controlled trials of movement therapies (R. P. Duncan & Earhart, 2012) and medication efficacy (Rascol et al., 2011). Given that changes in these scores are used to imply changes in motor severity, and consequently evaluate the efficacy of a given treatment, their lack of measurement accuracy may cast doubt on these conclusions.

7.2.3 Statistical Methods for Subtyping

The key implications of the present thesis concern the application and interpretation of subtyping analyses in the PD literature. The current approaches, such as K-means (hierarchical cluster analysis) and latent profile analysis, make strong assumptions about the relationships between observed measurements (Magidson & Vermunt, 2002a). As previously discussed, these analyses depend on the assumption of local independence (i.e. no relationships between variables within each subtype). The Bayesian analyses employed in Study 3 however, identified relationships between motor symptoms that were different in each of the extracted subtypes.

There are two primary implications of this finding. Firstly, these results show a clear utility for a Bayesian methodology in PD subtyping. The statistical approaches used in the seminal studies from Lewis et al. (2005; popularising K-means subtyping) and van Rooden et al. (2011; introducing latent profile analysis), and subsequently adopted by the wider literature, are unable to capture these relationships. Whereas the current standards for PD subtyping can identify which symptoms differ between groups, they are unable to describe how these differences present at varying levels of disease severity. Take, for example, the negative correlation that was found between left- and right-sided severity in the Asymmetric Undifferentiated (AU) subtype. K-means or LPA approaches to subtyping would not have been able to identify this relationship between motor symptom lateralities and as such would not have been able to provide this additional level of clinical description for the subtype. The methods proposed in the present thesis allow for this increased detail in PD subtyping.

Secondly, a Bayesian approach to PD subtyping also allows for the inclusion of far more precise measurements in the subtyping analysis. As discussed earlier, the use of sum or mean scores to represent motor symptom severity is likely to result in biased and unrepresentative measurements. van Rooden et al. (2011) attempted to circumvent this by estimating factor scores from a separate CFA. However, as with any estimation process, this introduces a small amount of error into the estimated scores (Grice, 2001). Of further concern, the CFA model used by van Rooden et al. (2011) did not account for the biases introduced by cross-loadings and residual covariances (van Rooden, Visser, Verbaan, Marinus, & van Hilten, 2009). Bias is highly likely to be introduced by this estimation of scores from an unrepresentative factor model. By incorporating a representative factor model into the subtyping process itself (as implemented in Study 3), rather than estimating the scores separately, individuals can be subtyped on the basis of representative and error-free measurements. However, as discussed in Chapter 4, this is only computationally possible in a Bayesian framework. Because the standard methods in PD subtyping have not been updated with the developments in statistical analyses, these methods are limited in both their precision and depth of exploration.

While the results of the present thesis do have negative implications for the reliability and precision of previous subtyping methods, of concern are the implications for the validity of previous results. Given that the subtyping analysis in the present thesis identified several relationships between the symptoms within each subtype, it is reasonable to suggest that this also would have occurred in previous subtyping studies and has not been accounted for. This is a clear violation of the assumption of local independence. For both K-means and LPA, violating the assumption of local independence can result in the over-extraction of clusters (i.e., more subtypes are identified than are actually present; Magidson & Vermunt, 2002a; Vermunt & Magidson, 2002). This implies a risk that previous subtyping studies that had not

accounted for relationships within subtypes may have identified more subtypes than are actually present. However, without evaluating the data collected by these studies, this risk is more hypothesis than actionable fact.

7.3 Recommendations for Researchers

The findings of the present thesis have emphasised the deficiencies that are present in current approaches to PD subtyping research. As the field progresses, there are three main areas of concern to be addressed by researchers: the symptoms/measures included in a subtyping analysis, the subtyping analysis itself, and the utility of Bayesian statistics.

7.3.1 Symptoms and Measures for Subtyping

The accuracy of PD research could benefit from greater emphasis on quality of measurement. The psychometric literature has seen significant developments in the past decade, with the introduction of methods such as exploratory structural equation modelling (ESEM; Asparouhov & Muthén, 2009), BSEM (Rindskopf, 2012), and approximate measurement invariance (Asparouhov & Muthen, 2014). Yet, aside from the present author's own work, measurement evaluation in PD is still conducted using either EFA or CFA (Corti et al., 2018; Johnson et al., 2016). While these methods may appear complex, Study 2 demonstrated that they can provide substantial improvements in the ability to describe a complex population. Consequently, it is recommended that PD researchers begin applying more recent psychometric methods when evaluating symptom measurement.

As a natural extension to the recommendations concerning symptom measurement, PD subtyping researchers should also reconsider the use of sum/mean scores where possible. Study 2 highlighted the limitations of this form of scoring in the presence of complex presentations (e.g., motor symptom measurement). The ideal alternative for subtyping is to incorporate both the measurement model (EFA/CFA/ESEM/BSEM) and the subtyping model

in the same analysis, as was undertaken in Study 3. It should be noted, however, that with many variables and multiple scales of measurement, this may only be computationally possible within a large enough sample size and the Bayesian framework presented in Study 3. Where this is not possible, factor scores can be a compromise between precision and practicality (Grice, 2001). While the estimation of factor scores will, in and of itself, introduce error, these scores will likely hold less measurement error than the corresponding sum or mean. However, it must be emphasised that factor scores will only be more reliable and precise if they are estimated from an accurate and well-fitting factor model. If a measurement model does not accurately represent the presentation of the disease, then the factor scores estimated from this model will also not be accurate. To date, only three subtyping studies have used factor scores in their analyses (Lawton et al., 2015; Liepelt-Scarfone et al., 2012; van Rooden et al., 2011). It is recommended that PD researchers begin to incorporate factor models or factor scores into their subtyping analyses to improve accuracy and reliability.

7.3.2 Subtyping Analyses

It would be simple enough to solely recommend that subtyping researchers use the proposed framework moving forward, and while this would be a marked improvement, the widespread use of k-means and LPA in PD are indicative of a greater concern. K-means and LPA have a broad range of applications and are powerful tools when used appropriately (Vermunt & Magidson, 2002). The present thesis, however, suggests that this may not be the case in PD. As the field progresses it is recommended that more attention is paid to the assumptions underlying statistical analyses, and how these reflect assumptions about the nature of symptoms under study. Given the complexity of PD, care needs to be taken to avoid the restrictive assumptions present in many analyses (such as the assumption of local independence).

That is not to say that the proposed framework of Study 3 should be ignored. The proposed methods provide a means of addressing these restrictive statistical assumptions while subtyping. Beyond this, the proposed method is the only subtyping model to date to account for measurement error in the analysis itself. PD subtyping researchers are recommended to apply the Bayesian subtyping approach presented in this thesis to improve the statistical power, reliability, and substantive accuracy of their analyses.

7.3.3 Utility of Bayesian Analyses

Finally, the possible benefits of Bayesian analyses for PD research as a whole should also be emphasised. These analyses allow for greater flexibility in model structure through the appropriate use of prior information (as demonstrated in Studies 2 and 3; Rindskopf, 2012). This flexibility will be important for the field moving forward, as it provides a means of addressing the previously discussed restrictive statistical assumptions. Bayesian analyses, when used appropriately, are also much more robust and powerful in smaller sample sizes (McNeish, 2016). Given the difficulties inherent to recruitment of a heterogenous clinical population, improved statistical power is of clear utility.

7.4 Recommendations for Clinicians

Whereas there were several methodological and statistical recommendations for researchers in PD, the recommendations for clinicians are more focused on the practical application of these research findings. For clinicians, there are three primary areas for application: differential rates of change in severity, accurate measurement of motor symptoms, and groupings of symptom presentation.

An improved understanding of how different Parkinsonian symptoms change in severity over time can inform a more efficient application of resources towards the management of a particular symptom. Rather than applying equal focus to each facet of the

disease presentation, treatments can be focused on the symptoms that are most likely to increase in severity. The present thesis identified that postural symptoms will increase in severity at a significantly faster rate than tremor symptoms. For a clinician, this would imply that, given equal severity, the patient would experience better outcomes from interventions targeting postural stability than tremor severity. Appropriate interventions could be programs such as tai chi (Li et al., 2012) or yoga (Sharma, Robbins, Wagner, & Colgrove, 2015). Unfortunately, given that the present thesis examined changes in two PD motor symptoms over time, the depth of clinical recommendation that can be made is limited.

The outcome of primary clinical benefit from this thesis is the ability to quantify motor symptoms, as well as symptom lateralities, using the provided calculator. Rather than assessing changes in each of the 33 items in the MDS-UPDRS motor examination, clinicians can now summarise the severity of each motor symptom for a given individual. The ability to quantify the severity of each laterality also provides clinicians to ability to assess for asymmetry in motor symptom presentation and track its change over time. This additional level of information in motor symptom presentation can allow clinicians to more accurately identify subtle changes in symptom severity over time.

While the presence of PD subtypes would not be unknown to clinicians, the inconsistencies in the literature have made clinical application difficult. Unfortunately, a complete solution to this is not possible in a single thesis. Without the consistent application of a single method in the literature it is difficult to translate subtyping findings to a clinical setting. However, the present findings do provide some guidance for clinicians. Given the findings of Studies 3 and 4, clinicians should be careful to monitor the quality of life of individuals with increased postural and rigidity symptoms. While it is not yet possible to say with certainty that these subtypes are present, these studies do indicate an increased likelihood of individuals with postural and rigidity symptoms experiencing poorer quality of

life in activities of daily living and communication. Until these methods can see repeated applications in PD subtyping research, the increased robustness at each stage of analyses should provide greater confidence in the results.

7.5 Closing Words

The idea that increased statistical complexity necessarily results in more ‘abstract’ or less clinically relevant findings is unfortunately common in the medical field. Too often there is a perception that statistical methods are distinct from clinical applications. The reality is, however, that the human experience is incredibly complex; there are few places this is more apparent than Parkinson’s disease. The sheer variety in individual experiences and disease outcomes has challenged the field’s understanding for over a hundred years. With this variety comes the need for more precise and more complex methods. One goal of this thesis was to demonstrate that more advanced statistics can be clinically relevant in Parkinson’s disease, that complex methods are needed to represent the complex experience of these individuals.

While this thesis has had a strong statistical and methodological focus, the emphasis has always remained on the relevance for the individual. People with Parkinson’s disease, and their loved ones, are faced with the prospect of living with an incurable disease. This is a very real burden and any research in this field should be completed with the aim of relieving, or at the very least, understanding this burden. The results and methods of this thesis are not presented out of academic curiosity or methodological novelty, but for the field of Parkinson’s research to better understand and help these individuals.

It is my hope that the work of this thesis provides at least a small step in progressing our understanding of Parkinson’s disease, and that this step brings us yet closer to easing the burden faced by these individuals.

“When nothing seems to help, I go and look at a stonecutter hammering away at his rock, perhaps a hundred times without as much as a crack showing in it. Yet at the hundred and first blow it will split in two, and I know it was not that last blow that did it, but all that had gone before.” - Jacob A. Riis

8 References

Aarsland, D., Andersen, K., Larsen, J. P., Lolk, A., & Kragh-Sorensen, P. (2003). Prevalence and characteristics of dementia in Parkinson disease: An 8-year prospective study. *Archives of Neurology, 60*(3), 387-392. doi:10.1001/archneur.60.3.387

Alves, G., Larsen, J. P., Emre, M., Wentzel-Larsen, T., & Aarsland, D. (2006). Changes in motor subtype and risk for incident dementia in Parkinson's disease. *Movement Disorders, 21*(8), 1123-1130. doi:10.1002/mds.20897

Alves, G., Wentzel-Larsen, T., Aarsland, D., & Larsen, J. P. (2005). Progression of motor impairment and disability in Parkinson disease: A population-based study. *Neurology, 65*(9), 1436-1441. doi:10.1212/01.wnl.0000183359.50822.f2

Andreadou, E., Anagnostouli, M., Vasdekis, V., Kararizou, E., Rentzos, M., Kontaxis, T., & Evdokimidis, I. (2011). The impact of comorbidity and other clinical and sociodemographic factors on health-related quality of life in Greek patients with Parkinson's disease. *Aging & Mental Health, 15*(7), 913-921. doi:10.1080/13607863.2011.569477

Antonini, A., Abbruzzese, G., Ferini-Strambi, L., Tilley, B., Huang, J., Stebbins, G. T., . . . Del Sorbo, F. (2013). Validation of the Italian version of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale. *Neurological Sciences, 34*(5), 683-687. doi:10.1007/s10072-012-1112-z

Asparouhov, T., & Muthen, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508. doi:10.1080/10705511.2014.919210

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438. doi:10.1080/10705510903008204

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation.

Asparouhov, T., & Muthén, B. (2011). *Using Bayesian priors for more flexible latent class analysis.* Paper presented at the Proceedings of the 2011 Joint Statistical Meeting, Miami Beach, FL.

Asparouhov, T., & Muthén, B. (2012). *Comparison of computational methods for high dimensional item factor analysis*. Retrieved from https://www.statmodel.com/download/HighDimension.pdf

Baba, T., Kikuchi, A., Hirayama, K., Nishio, Y., Hosokai, Y., Kanno, S., . . . Takeda, A. (2012). Severe olfactory dysfunction is a prodromal symptom of dementia associated with Parkinson's disease: A 3 year longitudinal study. *Brain, 135*, 161-169. doi:10.1093/brain/awr321

Bartolic, A., Pirtosek, Z., Rozman, J., & Ribaric, S. (2005). Postural stability of Parkinson's disease patients is improved by decreasing rigidity. *European Journal of Neurology, 12*(2), 156-159. doi:10.1111/j.1468-1331.2004.00942.x

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Beretta, V. S., Vitorio, R., Santos, P., Orcioli-Silva, D., & Gobbi, L. T. B. (2019). Postural control after unexpected external perturbation: Effects of Parkinson's disease subtype. *Human Movement Science, 64*, 12-18. doi:10.1016/j.humov.2019.01.001

Berlin, K. S., Williams, N. A., & Parra, G. R. (2013). An introduction to latent variable mixture modeling (Part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology, 39*(2), 174-187. doi:10.1093/jpepsy/jst084

Bohnen, N. I., & Albin, R. L. (2011). The cholinergic system and Parkinson disease. *Behavioural Brain Research, 221*(2), 564-573. doi:10.1016/j.bbr.2009.12.048

Braak, H., Tredici, K. D., Rüb, U., de Vos, R. A. I., Jansen Steur, E. N. H., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging, 24*(2), 197-211. doi:10.1016/s0197-4580(02)00065-9

Brissaud, E. (1925). *Leçons sur les maladies nerveuses*. Masson, Paris.

Broussolle, E., Krack, P., Thobois, S., Xie-Brustolin, J., Pollak, P., & Goetz, C. G. (2007). Contribution of Jules Froment to the study of Parkinsonian rigidity. *Movement Disorders, 22*(7), 909-914. doi:10.1002/mds.21484

Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York, United States: Guilford Publications.

Burn, D. J., Rowan, E. N., Allan, L. M., Molloy, S., O'Brien, J. T., & McKeith, I. G. (2006). Motor subtype and cognitive decline in Parkinson's disease, Parkinson's disease with dementia, and dementia with Lewy bodies. *Journal of Neurology, Neurosurgery and Psychiatry, 77*(5), 585-589. doi:10.1136/jnnp.2005.081711

Caap-Ahlgren, M., & Lannerheim, L. (2002). Older Swedish women's experiences of living with symptoms related to Parkinson's disease. *Journal of Advanced Nursing, 39*(1), 87-95. doi:10.1046/j.1365-2648.2002.02245.x

Cantiniaux, S., Vaugoyeau, M., Robert, D., Horrelou-Pitek, C., Mancini, J., Witjas, T., & Azulay, J. P. (2010). Comparative analysis of gait and speech in Parkinson's disease: hypokinetic or dysrhythmic disorders? *Journal of Neurology, Neurosurgery and Psychiatry, 81*(2), 177-184. doi:10.1136/jnnp.2009.174375

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1-29. doi:10.18637/jss.v076.i01

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195-212. doi:10.1007/Bf01246098

Corti, E. J., Johnson, A. R., Gasson, N., Bucks, R. S., Thomas, M. G., & Loftus, A. M. (2018). Factor structure of the Ways of Coping Questionnaire in Parkinson's disease. *Parkinsons Disease, 2018*. doi:10.1155/2018/7128069

Damholdt, M. F., Shevlin, M., Borghammer, P., Larsen, L., & Ostergaard, K. (2012). Clinical heterogeneity in Parkinson's disease revisited: A latent profile analysis. *Acta Neurologica Scandinavica, 125*(5), 311-318. doi:10.1111/j.1600-0404.2011.01561.x

Dauwerse, L., Hendrikx, A., Schipper, K., Struiksma, C., & Abma, T. A. (2014). Quality-of-life of patients with Parkinson's disease. *Brain Injury, 28*(10), 1342-1352. doi:10.3109/02699052.2014.916417

Davie, C. A. (2008). A review of Parkinson's disease. *British Medical Bulletin, 86*(1), 109-127. doi:10.1093/bmb/ldn013

Deuschl, G., Raethjen, J., Baron, R., Lindemann, M., Wilms, H., & Krack, P. (2000). The pathophysiology of Parkinsonian tremor: A review. *Journal of Neurology, 247*(5), V33-48. doi:10.1007/pl00007781

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

Djaldetti, R., Ziv, I., & Melamed, E. (2006). The mystery of motor asymmetry in Parkinson's disease. *Lancet Neurology, 5*(9), 796-802. doi:10.1016/S1474-4422(06)70549-X

Duncan, G. W., Khoo, T. K., Yarnall, A. J., O'Brien, J. T., Coleman, S. Y., Brooks, D. J., . . . Burn, D. J. (2014). Health-related quality of life in early Parkinson's disease: The impact of nonmotor symptoms. *Movement Disorders, 29*(2), 195-202. doi:10.1002/mds.25664

Duncan, R. P., & Earhart, G. M. (2012). Randomized controlled trial of community-based dancing to modify disease progression in Parkinson disease. *Neurorehabilitation and Neural Repair, 26*(2), 132-143. doi:10.1177/1545968311421614

Eggers, C., Kahraman, D., Fink, G. R., Schmidt, M., & Timmermann, L. (2011). Akinetic-rigid and tremor-dominant Parkinson's disease patients show different patterns of FP-CIT single photon emission computed tomography. *Movement Disorders, 26*(3), 416-423. doi:10.1002/mds.23468

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabil Psychol, 56*(4), 267-288. doi:10.1037/a0025579

Fahn, S., Elton, R. L., & UPDRS program members. (1987). Unified Parkinsons Disease Rating Scale. In S. Fahn, C. D. Marsden, M. Goldstein, & D. B. Calne (Eds.), *Recent Developments in Parkinsons Disease* (Vol. 2, pp. 153–163). Florham Park, NJ: Macmillan Healthcare Information.

Fereshtehnejad, S. M., Romenets, S. R., Anang, J. B., Latreille, V., Gagnon, J. F., & Postuma, R. B. (2015). New clinical subtypes of Parkinson disease and their longitudinal progression: A prospective cohort comparison with other phenotypes. *JAMA Neurology, 72*(8), 863-873. doi:10.1001/jamaneurol.2015.0703

Ferraye, M. U., Debu, B., Fraix, V., Goetz, L., Ardouin, C., Yelnik, J., . . . Pollak, P. (2010). Effects of pedunculopontine nucleus area stimulation on gait disorders in Parkinson's disease. *Brain, 133*, 205-214. doi:10.1093/brain/awp229

Foltynie, T., Brayne, C. E., Robbins, T. W., & Barker, R. A. (2004). The cognitive ability of an incident cohort of Parkinson's patients in the UK. The CamPaIGN study. *Brain, 127*, 550-560. doi:10.1093/brain/awh067

Ghiglione, P., Mutani, R., & Chio, A. (2005). Cogwheel rigidity. *Archives of Neurology, 62*(5), 828-830. doi:10.1001/archneur.62.5.828

Goetz, C. G. (2011). The history of Parkinson's disease: Early clinical descriptions and neurological therapies. *Cold Spring Harbor Perspectives in Medicine, 1*(1), a008862. doi:10.1101/cshperspect.a008862

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., . . . Movement Disorder Society, U. R. T. F. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders, 23*(15), 2129-2170. doi:10.1002/mds.22340

Graham, J. M., & Sagar, H. J. (1999). A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: Identification of three distinct subtypes. *Movement Disorders, 14*(1), 10-20. doi:10.1002/1531-8257(199901)14:1<10::AID-MDS1005>3.0.CO;2-4

Gregory, A. M., Buysse, D. J., Willis, T. A., Rijsdijk, F. V., Maughan, B., Rowe, R., . . . Eley, T. C. (2011). Associations between sleep quality and anxiety and depression symptoms in a sample of young adult twins and siblings. *Journal of Psychosomatic Research, 71*(4), 250-255. doi:10.1016/j.jpsychores.2011.03.011

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430-450. doi:10.1037/1082-989X.6.4.430

Gronau, Q. F., & Wagenmakers, E.-J. (2019). Limitations of Bayesian Leave-One-Out Cross-Validation for Model Selection. *Computational Brain & Behavior, 2*(1), 1-11. doi:10.1007/s42113-018-0011-7

Guan, X., Zeng, Q., Guo, T., Wang, J., Xuan, M., Gu, Q., . . . Zhang, M. (2017). Disrupted functional connectivity of basal ganglia across tremor-dominant and akinetic/rigid-dominant Parkinson's disease. *Frontiers in Aging Neuroscience, 9*(360), 360. doi:10.3389/fnagi.2017.00360

Halliday, G. M., & McCann, H. (2010). The progression of pathology in Parkinson's disease. *Annals of the New York Academy of Sciences, 1184*(1), 188-195. doi:10.1111/j.1749-6632.2009.05118.x

Hanley, J. A. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology, 157*(4), 364-375. doi:10.1093/aje/kwf215

Hariz, G. M., & Forsgren, L. (2011). Activities of daily living and quality of life in persons with newly diagnosed Parkinson's disease according to subtype of disease, and in comparison to healthy controls. *Acta Neurologica Scandinavica, 123*(1), 20-27. doi:10.1111/j.1600-0404.2010.01344.x

Helmich, R. C., Janssen, M. J., Oyen, W. J., Bloem, B. R., & Toni, I. (2011). Pallidal dysfunction drives a cerebellothalamic circuit into Parkinson tremor. *Annals of Neurology, 69*(2), 269-281. doi:10.1002/ana.22361

Hodaie, M., Neimat, J. S., & Lozano, A. M. (2007). The dopaminergic nigrostriatal system and Parkinson's disease: Molecular events in development, disease, and cell death, and new therapeutic strategies. *Neurosurgery, 60*(1), 17-28; discussion 28-30. doi:10.1227/01.NEU.0000249209.11967.CB

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression and mortality. *Neurology, 17*(5), 427-442. doi:10.1212/WNL.17.5.427

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery and Psychiatry, 55*(3), 181-184. doi:10.1136/jnnp.55.3.181

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651-666. doi:10.1016/j.patrec.2009.09.011

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology, Neurosurgery and Psychiatry, 79*(4), 368-376. doi:10.1136/jnnp.2007.131045

Jankovic, J., & Frost, J. D., Jr. (1981). Quantitative assessment of Parkinsonian and essential tremor: Clinical application of triaxial accelerometry. *Neurology, 31*(10), 1235-1240. doi:10.1212/WNL.31.10.1235

Jankovic, J., McDermott, M., Carter, J., Gauthier, S., Goetz, C., Golbe, L., . . . et al. (1990). Variable expression of Parkinson's disease: A base-line analysis of the DATATOP cohort. *Neurology, 40*(10), 1529-1534. doi:10.1212/wnl.40.10.1529

Jellinger, K. A. (2009). A critical evaluation of current staging of alpha-synuclein pathology in Lewy body disorders. *Biochimica et Biophysica Acta, 1792*(7), 730-740. doi:10.1016/j.bbadis.2008.07.006

Jenkinson, C., Fitzpatrick, R., Peto, V., Greenhall, R., & Hyman, N. (1997). The Parkinson's Disease Questionnaire (PDQ-39): Development and validation of a Parkinson's disease summary index score. *Age and Ageing, 26*(5), 353-357. doi:10.1093/ageing/26.5.353

Jenkinson, N., Nandi, D., Muthusamy, K., Ray, N. J., Gregory, R., Stein, J. F., & Aziz, T. Z. (2009). Anatomy, physiology, and pathophysiology of the pedunculopontine nucleus. *Movement Disorders, 24*(3), 319-328. doi:10.1002/mds.22189

Johns, P. (2014). Parkinson's Disease. In *Clinical Neuroscience*. London: Elsevier Health Sciences.

Johnson, A. R., Lawrence, B. J., Corti, E. J., Booth, L., Gasson, N., Thomas, M. G., . . . Bucks, R. S. (2016). Suitability of the Depression, Anxiety, and Stress Scale in Parkinson's Disease. *Journal of Parkinsons Disease, 6*(3), 609-616. doi:10.3233/JPD-160842

Kehagia, A. A., Barker, R. A., & Robbins, T. W. (2010). Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. *Lancet Neurology, 9*(12), 1200-1213. doi:10.1016/S1474-4422(10)70212-X

Kim, J. S., Yang, J. J., Lee, J. M., Youn, J., Kim, J. M., & Cho, J. W. (2014). Topographic pattern of cortical thinning with consideration of motor laterality in Parkinson disease. *Parkinsonism & Related Disorders, 20*(11), 1186-1190. doi:10.1016/j.parkreldis.2014.08.021

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

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178-206. doi:10.3758/s13423-016-1221-4

LaBelle, D. R., Walsh, R. R., & Banks, S. J. (2017). Latent cognitive phenotypes in de novo Parkinson's disease: A person-centered approach. *Journal of the International Neuropsychological Society, 23*(7), 551-563. doi:10.1017/S1355617717000406

Lanza, S. T., Bray, B. C., & Collins, L. M. (2013). An introduction to latent class and latent transition analysis. In *Handbook of psychology: Research methods in psychology, Vol. 2, 2nd ed.* (pp. 691-716). Hoboken, NJ, US: John Wiley & Sons Inc.

Lawton, M., Baig, F., Rolinski, M., Ruffman, C., Nithi, K., May, M. T., . . . Hu, M. T. (2015). Parkinson's disease subtypes in the Oxford Parkinson Disease Centre (OPDC) discovery cohort. *Journal of Parkinsons Disease, 5*(2), 269-279. doi:10.3233/JPD-140523

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989-2001. doi:10.1016/j.jmva.2009.04.008

Lewis, S. J., Foltynie, T., Blackwell, A. D., Robbins, T. W., Owen, A. M., & Barker, R. A. (2005). Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *Journal of Neurology, Neurosurgery and Psychiatry, 76*(3), 343-348. doi:10.1136/jnnp.2003.033530

Li, F., Harmer, P., Fitzgerald, K., Eckstrom, E., Stock, R., Galver, J., . . . Batya, S. S. (2012). Tai chi and postural stability in patients with Parkinson's disease. *New England Journal of Medicine, 366*(6), 511-519. doi:10.1056/NEJMoa1107911

Liepelt-Scarfone, I., Graber, S., Fruhmann Berger, M., Feseker, A., Baysal, G., Csoti, I., . . . Berg, D. (2012). Cognitive profiles in Parkinson's disease and their relation to dementia: A data-driven approach. *International Journal of Alzheimer's Disease, 2012*, 910757. doi:10.1155/2012/910757

Lindstrom, J. C., Wyller, N. G., Halvorsen, M. M., Hartberg, S., & Lundqvist, C. (2017). Psychometric properties of a Norwegian adaption of the Barratt Impulsiveness Scale-11 in a sample of Parkinson patients, headache patients, and controls. *Brain & Behavior, 7*(1), e00605. doi:10.1002/brb3.605

Litvan, I., Aarsland, D., Adler, C. H., Goldman, J. G., Kulisevsky, J., Mollenhauer, B., . . . Weintraub, D. (2011). MDS Task Force on mild cognitive impairment in Parkinson's disease: Critical review of PD-MCI. *Movement Disorders, 26*(10), 1814-1824. doi:10.1002/mds.23823

Litvan, I., Goldman, J. G., Troster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., . . . Emre, M. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's

disease: Movement Disorder Society Task Force guidelines. *Movement Disorders, 27*(3), 349-356. doi:10.1002/mds.24893

Liu, J. B., Leng, J. L., Wang, Y. G., Zhang, Y., Tang, T. Y., Tao, L. H., . . . Liu, C. F. (2019). Investigation of nonmotor symptoms in first-degree relatives of patients with different clinical types of Parkinson's disease. *Parkinsons Disease, 2019*, 1654161. doi:10.1155/2019/1654161

Liu, P., Feng, T., Wang, Y. J., Zhang, X., & Chen, B. (2011). Clinical heterogeneity in patients with early-stage Parkinson's disease: A cluster analysis. *Journal of Zhejiang University SCIENCE B, 12*(9), 694-703. doi:10.1631/jzus.B1100069

Loftus, A. M., Bucks, R. S., Thomas, M., Kane, R., Timms, C., Barker, R. A., & Gasson, N. (2015). Retrospective assessment of Movement Disorder Society criteria for mild cognitive impairment in Parkinson's disease. *Journal of the International Neuropsychological Society, 21*(2), 137-145. doi:10.1017/S1355617715000041

Lord, S., Galna, B., Coleman, S., Yarnall, A., Burn, D., & Rochester, L. (2014). Cognition and gait show a selective pattern of association dominated by phenotype in incident Parkinson's disease. *Frontiers in Aging Neuroscience, 6*, 249. doi:10.3389/fnagi.2014.00249

Louis, E. D., Tang, M. X., Cote, L., Alfaro, B., Mejia, H., & Marder, K. (1999). Progression of Parkinsonian signs in Parkinson disease. *Archives of Neurology, 56*(3), 334-337. doi:10.1001/archneur.56.3.334

Lubke, G. H., & Luningham, J. (2017). Fitting latent variable mixture models. *Behaviour Research and Therapy, 98*, 91-102. doi:10.1016/j.brat.2017.04.003

Lubke, G. H., & Muthen, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 21-39. doi:10.1037/1082-989X.10.1.21

Lundervold, A. J., Karlsen, N. R., & Reinvang, I. (1994). Assessment of 'subcortical dementia' in patients with Huntington's disease, Parkinson's disease, multiple sclerosis and AIDS by a neuropsychological screening battery. *Scandinavian Journal of Psychology, 35*(1), 48-55. doi:10.1111/j.1467-9450.1994.tb00932.x

Lunn, D., Jackson, C., Best, N., Spiegelhalter, D., & Thomas, A. (2012). *The BUGS book: A practical introduction to Bayesian analysis*: Chapman and Hall/CRC.

Magidson, J., & Vermunt, J. K. (2002a). Latent class modeling as a probabilistic extension of K-means clustering. *Quirk's Marketing Research Review, 20*(3), 77-80.

Magidson, J., & Vermunt, J. K. (2002b). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research, 20*(1), 36-43.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., . . . Taylor, P. (2011). The Parkinson Progression Marker Initiative (PPMI). *Progress in Neurobiology, 95*(4), 629-635. doi:10.1016/j.pneurobio.2011.09.005

Marras, C., & Lang, A. (2013). Parkinson's disease subtypes: Lost in translation? *Journal of Neurology, Neurosurgery and Psychiatry, 84*(4), 409-415. doi:10.1136/jnnp-2012-303455

Martinez-Martin, P., Jeukens-Visser, M., Lyons, K. E., Rodriguez-Blazquez, C., Selai, C., Siderowf, A., . . . Schrag, A. (2011). Health-related quality-of-life scales in Parkinson's disease: Critique and recommendations. *Movement Disorders, 26*(13), 2371-2380. doi:10.1002/mds.23834

Martinez-Martin, P., Rodriguez-Blazquez, C., Alvarez-Sanchez, M., Arakaki, T., Bergareche-Yarza, A., Chade, A., . . . Goetz, C. G. (2013). Expanded and independent validation of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). *Journal of Neurology, 260*(1), 228-236. doi:10.1007/s00415-012-6624-1

McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research, 18*(3-4), 145-166. doi:10.1080/03610739208253917

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750-773. doi:10.1080/10705511.2016.1186549

Merchant, H., Luciana, M., Hooper, C., Majestic, S., & Tuite, P. (2008). Interval timing and Parkinson's disease: Heterogeneity in temporal performance. *Experimental Brain Research, 184*(2), 233-248. doi:10.1007/s00221-007-1097-7

Micieli, G., Tosi, P., Marcheselli, S., & Cavallini, A. (2003). Autonomic dysfunction in Parkinson's disease. *Neurological Sciences, 24 Suppl 1*(1), S32-34. doi:10.1007/s100720300035

Mîndrilă, D. (2010). Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal for Digital Society, 1*(1), 60-66. doi:10.20533/ijds.2040.2570.2010.0010

Morris, M., Iansek, R., Smithson, F., & Huxham, F. (2000). Postural instability in Parkinson's disease: A comparison with and without a concurrent task. *Gait and Posture, 12*(3), 205-216. doi:Doi 10.1016/S0966-6362(00)00076-X

Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. (2003). The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders, 18*(7), 738-750. doi:10.1002/mds.10473

Muthen, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313-335. doi:10.1037/a0026802

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*(1), 190-204. doi:10.1006/jmps.1999.1283

Nabli, F., Ben Sassi, S., Amouri, R., Duda, J. E., Farrer, M. J., & Hentati, F. (2015). Motor phenotype of LRRK2-associated Parkinson's disease: A Tunisian longitudinal study. *Movement Disorders, 30*(2), 253-258. doi:10.1002/mds.26097

Neely-Barnes, S. (2010). Latent class models in social work. *Social Work Research, 34*(2), 114-121. doi:10.1093/swr/34.2.114

Nutt, J. G. (2016). Motor subtype in Parkinson's disease: Different disorders or different stages of disease? *Movement Disorders, 31*(7), 957-961. doi:10.1002/mds.26657

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson & M. Kaptein (Eds.), *Modern Statistical Methods for HCI* (pp. 275-287). Cham: Springer International Publishing.

Olanow, C. W., & Brundin, P. (2013). Parkinson's disease and alpha synuclein: Is Parkinson's disease a prion-like disorder? *Movement Disorders, 28*(1), 31-40. doi:10.1002/mds.25373

Pahapill, P. A., & Lozano, A. M. (2000). The pedunculopontine nucleus and Parkinson's disease. *Brain, 123 (Pt 9)*(9), 1767-1783. doi:10.1093/brain/123.9.1767

Pavese, N., Rivero-Bosch, M., Lewis, S. J., Whone, A. L., & Brooks, D. J. (2011). Progression of monoaminergic dysfunction in Parkinson's disease: A longitudinal 18F-dopa PET study. *Neuroimage, 56*(3), 1463-1468. doi:10.1016/j.neuroimage.2011.03.012

Pelicioni, P. H. S., Brodie, M. A., Latt, M. D., Menant, J. C., Menz, H. B., Fung, V. S. C., & Lord, S. R. (2018). Head and trunk stability during gait before and after levodopa intake in Parkinson's disease subtypes. *Experimental Gerontology, 111*, 78-85. doi:10.1016/j.exger.2018.06.031

Peto, V., Jenkinson, C., & Fitzpatrick, R. (1998). PDQ-39: A review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. *Journal of Neurology, 245*(S1), S10-S14. doi:10.1007/pl00007730

Pinter, D., Kovacs, M., Harmat, M., Juhasz, A., Janszky, J., & Kovacs, N. (2019). Trimetazidine and Parkinsonism: A prospective study. *Parkinsonism & Related Disorders*. doi:10.1016/j.parkreldis.2019.01.005

Pirker, W., Djamshidian, S., Asenbaum, S., Gerschlager, W., Tribl, G., Hoffmann, M., & Brucke, T. (2002). Progression of dopaminergic degeneration in Parkinson's disease and atypical parkinsonism: A longitudinal beta-CIT SPECT study. *Movement Disorders, 17*(1), 45-53. doi:10.1002/mds.1265

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics, 9*(3), 523-539. doi:10.1093/biostatistics/kxm049

Politis, M., Wu, K., Molloy, S., P, G. B., Chaudhuri, K. R., & Piccini, P. (2010). Parkinson's disease symptoms: The patient's perspective. *Movement Disorders, 25*(11), 1646-1651. doi:10.1002/mds.23135

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., . . . Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders, 30*(12), 1591-1601. doi:10.1002/mds.26424

Pushpanathan, M. E., Loftus, A. M., Gasson, N., Thomas, M. G., Timms, C. F., Olaithe, M., & Bucks, R. S. (2018). Beyond factor analysis: Multidimensionality and the Parkinson's Disease Sleep Scale-Revised. *PLOS ONE, 13*(2), e0192394. doi:10.1371/journal.pone.0192394

Pushpanathan, M. E., Loftus, A. M., Thomas, M. G., Gasson, N., & Bucks, R. S. (2016). The relationship between sleep and cognition in Parkinson's disease: A meta-analysis. *Sleep Medicine Reviews, 26*, 21-32. doi:10.1016/j.smrv.2015.04.003

Rajput, A. H., Sitte, H. H., Rajput, A., Fenton, M. E., Pifl, C., & Hornykiewicz, O. (2008). Globus pallidus dopamine and Parkinson motor subtypes. *Neurology, 70*(16 Part 2), 1403. doi:10.1212/01.wnl.0000285082.18969.3a

Ramsawh, H. J., Stein, M. B., Belik, S. L., Jacobi, F., & Sareen, J. (2009). Relationship of anxiety disorders, sleep quality, and functional impairment in a community sample. *Journal of Psychiatric Research, 43*(10), 926-933. doi:10.1016/j.jpsychires.2009.01.009

Rascol, O., Fitzer-Attas, C. J., Hauser, R., Jankovic, J., Lang, A., Langston, J. W., . . . Olanow, C. W. (2011). A double-blind, delayed-start trial of rasagiline in Parkinson's disease (the ADAGIO study): Prespecified and post-hoc analyses of the need for additional therapies, changes in UPDRS scores, and non-motor outcomes. *Lancet Neurology, 10*(5), 415-423. doi:10.1016/S1474-4422(11)70073-4

Reijnders, J. S., Ehrt, U., Lousberg, R., Aarsland, D., & Leentjens, A. F. (2009). The association between motor subtypes and psychopathology in Parkinson's disease. *Parkinsonism & Related Disorders, 15*(5), 379-382. doi:10.1016/j.parkreldis.2008.09.003

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16 Suppl 1*(1), 19-31. doi:10.1007/s11136-007-9183-7

Rindskopf, D. (2012). Next steps in Bayesian structural equation models: Comments on, variations of, and extensions to Muthen and Asparouhov (2012). *Psychological Methods, 17*(3), 336-339; discussion 346-353. doi:10.1037/a0027130

Ross, K. A., Jensen, C. S., Snodgrass, R., Dyreson, C. E., Jensen, C. S., Snodgrass, R., . . . Chen, L. (2009). Cross-validation. In L. Liu & M. T. ÖZsu (Eds.), *Encyclopedia of Database Systems* (pp. 532-538). Boston, MA: Springer US.

Schiess, M. C., Zheng, H., Soukup, V. M., Bonnen, J. G., & Nauta, H. J. (2000). Parkinson's disease subtypes: Clinical classification and ventricular cerebrospinal fluid analysis. *Parkinsonism & Related Disorders, 6*(2), 69-76. doi:10.1016/S1353-8020(99)00051-6

Schrag, A., Jahanshahi, M., & Quinn, N. (2000). How does Parkinson's disease affect quality of life? A comparison with quality of life in the general population. *Movement Disorders, 15*(6), 1112-1118. doi:10.1002/1531-8257(200011)15:6<1112::Aid-mds1008>3.0.Co;2-a

Schrag, A., Quinn, N. P., & Ben-Shlomo, Y. (2006). Heterogeneity of Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 77*(2), 275-276.

Sharma, N. K., Robbins, K., Wagner, K., & Colgrove, Y. M. (2015). A randomized controlled pilot study of the therapeutic effects of yoga in people with Parkinson's disease. *International Journal of Yoga, 8*(1), 74-79. doi:10.4103/0973-6131.146070

Simuni, T., Caspell-Garcia, C., Coffey, C., Lasch, S., Tanner, C., Marek, K., & Investigators, P. (2016). How stable are Parkinson's disease subtypes in de novo patients: Analysis of the PPMI cohort? *Parkinsonism & Related Disorders, 28*, 62-67. doi:10.1016/j.parkreldis.2016.04.027

Singh, V. K., Tiwari, N., & Garg, S. (2011, 7-9 Oct. 2011). *Document Clustering Using K-Means, Heuristic K-Means and Fuzzy C-Means.* Paper presented at the 2011 International Conference on Computational Intelligence and Communication Networks.

Soh, S. E., Morris, M. E., & McGinley, J. L. (2011). Determinants of health-related quality of life in Parkinson's disease: A systematic review. *Parkinsonism & Related Disorders, 17*(1), 1-9. doi:10.1016/j.parkreldis.2010.08.012

Spillantini, M. G., & Goedert, M. (2000). The alpha-synucleinopathies: Parkinson's disease, dementia with Lewy bodies, and multiple system atrophy. *Annals of the New York Academy of Sciences, 920*(1), 16-27. doi:10.1111/j.1749-6632.2000.tb06900.x

Starkstein, S. E., Dragovic, M., Jorge, R., Brockman, S., Merello, M., Robinson, R. G., . . . Wilson, M. (2011). Diagnostic criteria for depression in Parkinson's disease: A study of symptom patterns using latent class analysis. *Movement Disorders, 26*(12), 2239-2245. doi:10.1002/mds.23836

Stebbins, G. T., Goetz, C. G., Burn, D. J., Jankovic, J., Khoo, T. K., & Tilley, B. C. (2013). How to identify tremor dominant and postural instability/gait difficulty groups with the Movement Disorder Society Unified Parkinson's Disease Rating Scale: Comparison with the Unified Parkinson's Disease Rating Scale. *Movement Disorders, 28*(5), 668-670. doi:10.1002/mds.25383

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893-898. doi:10.1016/j.paid.2006.09.017

Stochl, J., Boomsma, A., Ruzicka, E., Brozova, H., & Blahus, P. (2008). On the structure of motor symptoms of Parkinson's disease. *Movement Disorders, 23*(9), 1307-1312. doi:10.1002/mds.22029

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*(3), 1-16.

Sveinbjornsdottir, S. (2016). The clinical symptoms of Parkinson's disease. *Journal of Neurochemistry, 139 Suppl 1*(S1), 318-324. doi:10.1111/jnc.13691

Taylor Tavares, A. L., Jefferis, G. S., Koop, M., Hill, B. C., Hastie, T., Heit, G., & Bronte-Stewart, H. M. (2005). Quantitative measurements of alternating finger tapping in

Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Movement Disorders, 20*(10), 1286-1298. doi:10.1002/mds.20556

Thenganatt, M. A., & Jankovic, J. (2014). Parkinson disease subtypes. *JAMA Neurology, 71*(4), 499-504. doi:10.1001/jamaneurol.2013.6233

Tickle-Degnen, L., & Doyle Lyons, K. (2004). Practitioners' impressions of patients with Parkinson's disease: The social ecology of the expressive mask. *Social Science & Medicine, 58*(3), 603-614. doi:10.1016/s0277-9536(03)00213-2

Tolosa, E., Wenning, G., & Poewe, W. (2006). The diagnosis of Parkinson's disease. *Lancet Neurology, 5*(1), 75-86. doi:10.1016/S1474-4422(05)70285-4

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders, 25*(15), 2649-2653. doi:10.1002/mds.23429

Tubert, C., Galtieri, D., & Surmeier, D. J. (2018). The pedunclopontine nucleus and Parkinson's disease. *Neurobiological Disorders*. doi:10.1016/j.nbd.2018.08.017

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*(3), 842-860. doi:10.1111/cdev.12169

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology, 4*(770), 770. doi:10.3389/fpsyg.2013.00770

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492. doi:10.1080/17405629.2012.686740

van Dongen, S. (2006). Prior specification in Bayesian statistics: Three cautionary tales. *Journal of Theoretical Biology, 242*(1), 90-100. doi:10.1016/j.jtbi.2006.02.002

van Rooden, S. M., Colas, F., Martinez-Martin, P., Visser, M., Verbaan, D., Marinus, J., . . . van Hilten, J. J. (2011). Clinical subtypes of Parkinson's disease. *Movement Disorders, 26*(1), 51-58. doi:10.1002/mds.23346

van Rooden, S. M., Heiser, W. J., Kok, J. N., Verbaan, D., van Hilten, J. J., & Marinus, J. (2010). The identification of Parkinson's disease subtypes using cluster analysis: A systematic review. *Movement Disorders, 25*(8), 969-978. doi:10.1002/mds.23116

van Rooden, S. M., Visser, M., Verbaan, D., Marinus, J., & van Hilten, J. J. (2009). Motor patterns in Parkinson's disease: A data-driven approach. *Movement Disorders, 24*(7), 1042-1047. doi:10.1002/mds.22512

van Uem, J. M., Marinus, J., Canning, C., van Lummel, R., Dodel, R., Liepelt-Scarfone, I., . . . Maetzler, W. (2016). Health-related quality of life in patients with Parkinson's disease: A systematic review based on the ICF model. *Neuroscience Biobehavioural Review, 61*, 26-34. doi:10.1016/j.neubiorev.2015.11.014

Vehtari, A. (2017). Loo comparison in reference to standard error. Retrieved from http://discourse.mc-stan.org/t/loo-comparison-in-reference-to-standard-error/4009/3

Vehtari, A., Gelman, A., & Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413-1432. doi:10.1007/s11222-016-9696-4

Vermunt, J. K., & Magidson, J. (2002). Latent Class Cluster Analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis*. New York: Cambridge University Press.

Visser, M., Marinus, J., Stiggelbout, A. M., & Van Hilten, J. J. (2004). Assessment of autonomic dysfunction in Parkinson's disease: The SCOPA-AUT. *Movement Disorders, 19*(11), 1306-1312. doi:10.1002/mds.20153

Vu, T. C., Nutt, J. G., & Holford, N. H. (2012). Progression of motor and nonmotor features of Parkinson's disease and their response to treatment. *British Journal of Clinical Pharmacology, 74*(2), 267-283. doi:10.1111/j.1365-2125.2012.04192.x

Weintraub, D., & Burn, D. J. (2011). Parkinson's disease: The quintessential neuropsychiatric disorder. *Movement Disorders, 26*(6), 1022-1031. doi:10.1002/mds.23664

Weintraub, D., Koester, J., Potenza, M. N., Siderowf, A. D., Stacy, M., Voon, V., . . . Lang, A. E. (2010). Impulse control disorders in Parkinson disease: A cross-sectional study of 3090 patients. *Archives of Neurology, 67*(5), 589-595. doi:10.1001/archneurol.2010.65

Weintraub, D., Moberg, P. J., Duda, J. E., Katz, I. R., & Stern, M. B. (2004). Effect of psychiatric and other nonmotor symptoms on disability in Parkinson's disease. *Journal of the American Geriatric Society, 52*(5), 784-788. doi:10.1111/j.1532-5415.2004.52219.x

Weintraub, D., & Stern, M. B. (2005). Psychiatric complications in Parkinson disease. *American Journal of Geriatric Psychiatry, 13*(10), 844-851. doi:10.1176/appi.ajgp.13.10.844

White, N., Johnson, H., Silburn, P., Mellick, G., Dissanayaka, N., & Mengersen, K. (2012). Probabilistic subgroup identification using Bayesian finite mixture modelling: A case study in Parkinson's disease phenotype identification. *Statistical Methods in Medical Research, 21*(6), 563-583. doi:10.1177/0962280210391012

Yarnall, A., Rochester, L., & Burn, D. J. (2011). The interplay of cholinergic function, attention, and falls in Parkinson's disease. *Movement Disorders, 26*(14), 2496-2503. doi:10.1002/mds.23932

Zetusky, W. J., Jankovic, J., & Pirozzolo, F. J. (1985). The heterogeneity of Parkinson's disease: Clinical and prognostic implications. *Neurology, 35*(4), 522-526. doi:10.1212/wnl.35.4.522

Zhang, J., Wei, L., Hu, X., Xie, B., Zhang, Y., Wu, G. R., & Wang, J. (2015). Akinetic-rigid and tremor-dominant Parkinson's disease patients show different patterns of intrinsic brain activity. *Parkinsonism & Related Disorders, 21*(1), 23-30. doi:10.1016/j.parkreldis.2014.10.017

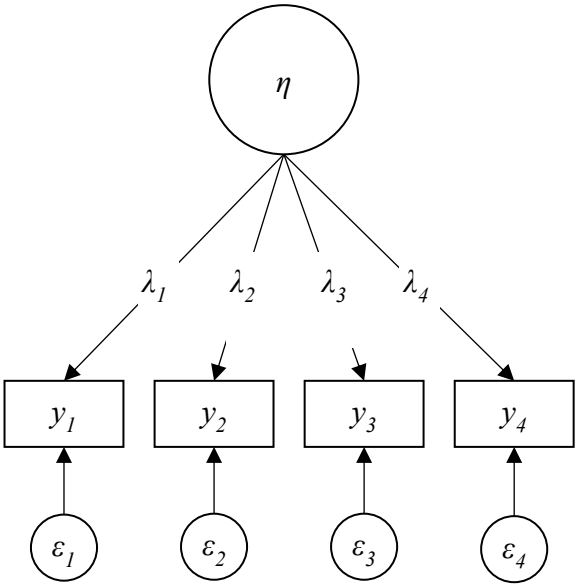
Zyphur, M. J., & Oswald, F. L. (2013). Bayesian estimation and inference. *Journal of Management, 41*(2), 390-420. doi:10.1177/0149206313501200

9 Appendices

9.1 Appendix A: Factor Model Parameterisation

9.1.1 Factor Analysis as Regression

Most will be familiar with the graphical depiction of a factor analysis, one in which an unobserved latent variable influences the level of the observed items (in this thesis, MDS-UPDRS measurements):



For those unfamiliar, the above diagram describes the idea of an unobserved latent variable η (a given motor symptom) affecting the observed severity of a given item y (MDS-UPDRS ratings). The factor loadings λ are the estimates of the strength of relationship between the severity of motor symptom η and the resultant rating on MDS-UPDRS item y. The residuals ε represent the elements other than motor symptoms that affect MDS-UPDRS item ratings (e.g. unrelated motor symptoms or antiparkinsonian medications).

We can also represent this factor model as a linear regression, as all the same components are present: there is the observed outcome y, covariate η, regression coefficient λ, and residual ε. Where many would be familiar with the form:

$$y = \beta x + e$$

We would instead write:

$$y = \lambda\eta + \varepsilon$$

9.1.2 Linear Regression as a Normal Distribution

While presenting factor analyses as a form of linear regression can help aid interpretation, it also simplifies the specification of these models for estimation. In a standard linear regression, the βx component represents the predicted mean score on item *y* for a given value of the covariate *x*. When presented graphically, βx is the 'line of best fit' through a scatterplot (i.e. the predicted value of *y* for each value of *x*). This predicted value is the same for all individuals with the same level of the covariate, but each will differ from this predicted score by a unique amount (i.e. their residual e). That is, the predicted mean βx is fixed (the same for each individual with the same *x*). In this way, a linear regression can also be represented as a normal distribution: the predicted value is the mean of the distribution, with the residuals representing the variation around this.

That is, instead of the regression equation:

$$y = \beta x + e$$

We can instead say that *y* is normally distributed with mean βx and variance e^2:

$$y \sim N(\beta x, e^2)$$

Equivalently, we can then specify a factor model as a normal distribution with mean $\lambda\eta$ and variance ε^2:

$$y \sim N(\lambda\eta, \varepsilon^2)$$

That is, λη represents an individual's predicted level of the outcome *y* for their level of the latent variable η, and ε represents the amount to which their observed level of the outcome *y* differs from this.

9.1.3 Categorical Ratings of Continuous Severity

However, while we can specify a factor analysis as a normal distribution, this is made somewhat more complex when our observed measurements are categorical. This is because categorical ratings cannot vary in the same way as continuous measurements. Using the MDS-UPDRS as an example: ratings can only take on 5 possible values to indicate the magnitude of severity. Contrast this with a measure such as cortical thickness, where the ratings can take on an infinite number of values depending on the precision of measurement (e.g. 2.50001mm or 2.50002mm). Because categorical ratings cannot vary in the same way as continuous measurements, they cannot capture the subtler variations in severity between individuals (in their current form, at least). Conceptually, it makes more sense to view the MDS-UPDRS rating categories as a means of grouping individuals with roughly similar levels of severity on a given assessment.

We can illustrate this using MDS-UPDRS item 3.18: Constancy of Rest Tremor as an example, the rating categories are:

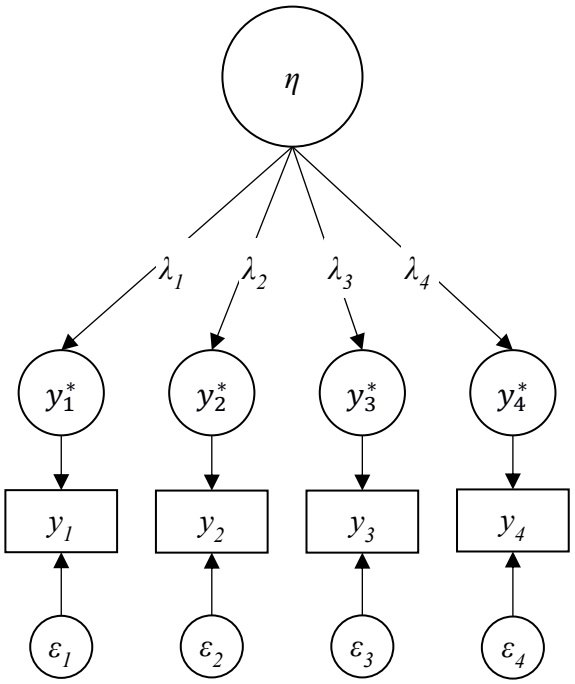
- Normal – No tremor
- Slight – Rest tremor present <25% of examination
- Mild – Rest tremor present 26-50% of examination
- Moderate – Rest tremor present 51-75% of examination
- Severe – Rest tremor present >75% of examination

While two individuals may have the same severity rating (e.g. 'Slight') it is likely that they would have exhibited rest tremors for slightly different proportions of the examination

(e.g. 10% vs. 12%). That is, regardless of an individual's specific level of rest tremor severity, their severity rating is determined by whether or not they exceed the threshold for each rating category.

9.1.4 Categorical Observations in Factor Analysis

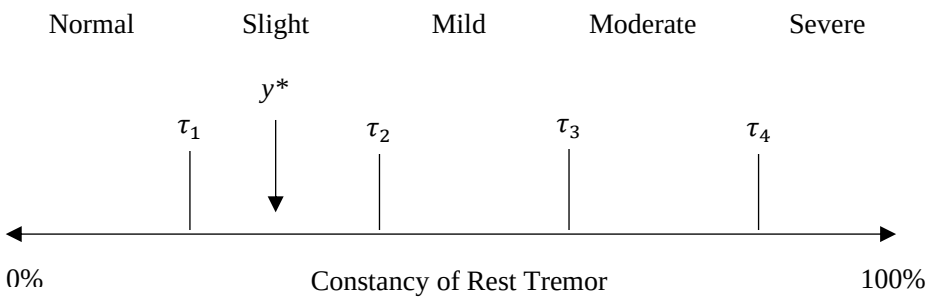
To represent this concept in a factor analytic context we introduce another latent variable (y*) for each observed measurement (y):



These latent y* now represent the specific levels of severity for each of the MDS-UPDRS items (e.g. 10% or 12%). We then estimate a set of thresholds (τ) for each item, which represents the point at which an individual's specific level of severity (y*) results in a different MDS-UPDRS severity rating. For example, the point at which an individual's severity changes from a rating of 'Slight' to a rating of 'Mild'.

This is depicted graphically below, where an individual's actual level of severity (y*) is larger than the upper threshold for the 'Normal' severity rating (τ_1), but is smaller than the upper threshold for the 'Slight' rating (τ_2), and so their observed rating (y) would be 'Slight'.

In clinical terms, the individual's symptom was too severe to be considered 'Normal', but not severe enough to be considered 'Mild'.



9.1.5 Categorical Factor Analysis as a Distribution

Earlier we covered the concept of a representing a factor model as a normal distribution, that is:

$$y \sim N(\lambda\eta, \varepsilon^2)$$

To extend this concept to factor analysis with categorical observations, we need to add an extra step. This is because the latent motor symptom variable η is no longer directly affecting the observed ratings of symptom severity (y). Instead, the latent motor symptom η is now influencing the individual's underlying, *continuous*, level of severity (y^*):

$$y^* = \lambda\eta + \varepsilon$$

Which we can then express as a normal distribution:

$$y^* \sim N(\lambda\eta, \varepsilon^2)$$

This continuous level of severity y^* is then related to the observed category ratings of severity (y) through the estimated thresholds τ:

$$y^* < \tau_1 = \text{Normal}$$

$$\tau_1 < y^* < \tau_2 = \text{Slight}$$

$$\tau_2 < y^* < \tau_3 = \text{Mild}$$

$$\tau_3 < y^* < \tau_4 = \text{Moderate}$$

$$\tau_4 < y^* = \text{Severe}$$

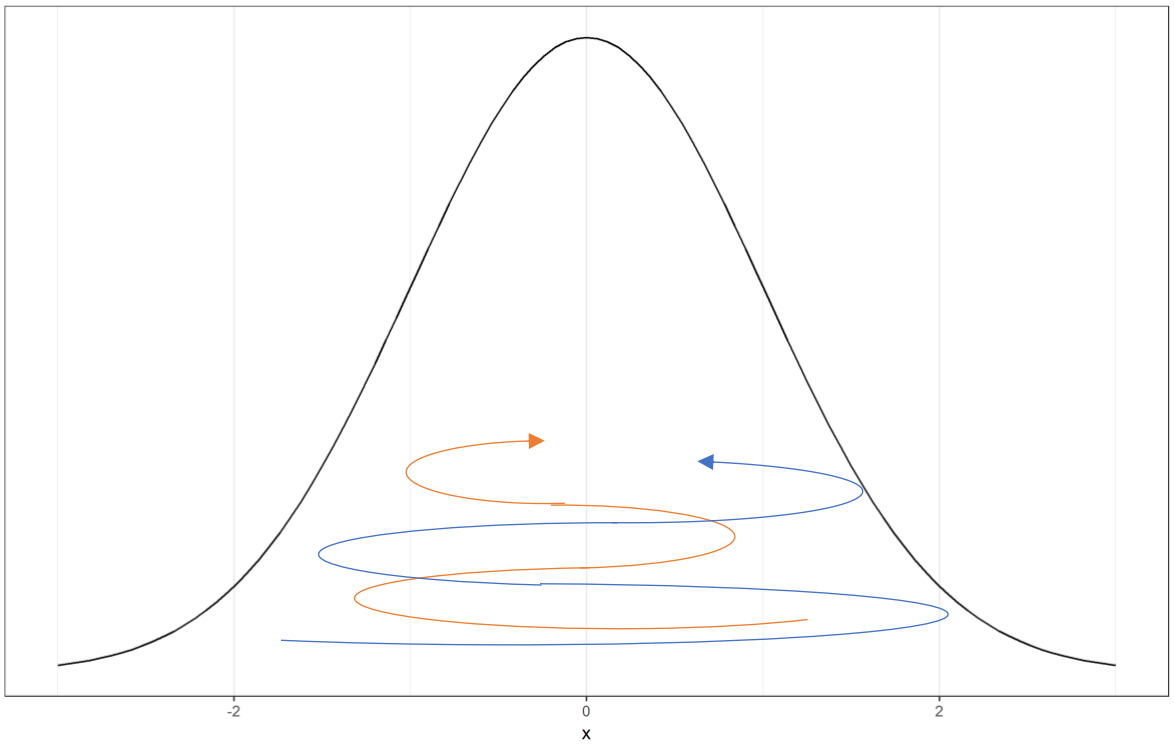
9.2 Appendix B: Difficulties with Bayesian Estimation of Factor Models

While it is (relatively) simple to specify the form of a factor analysis for a Bayesian sampler, there are still difficulties unique to Bayesian estimation that require careful consideration. To better understand these, it will be helpful to briefly review how estimation using these samplers differs from traditional approaches (primarily maximum-likelihood).

9.2.1 How Bayesian Sampling Differs

Maximum-likelihood can generally be considered an 'optimisation' procedure. Given initial values for each parameter in the model, and a function assessing fit to the data (the log-likelihood function), maximum-likelihood returns the parameter values that give the best fit. That is, they aim to maximise the likelihood function for a given model.

In contrast, Bayesian samplers aim to 'explore' the possible values that each parameter is likely to take. Take as an example the density plot pictured below. As a rough analogue of the approach of a Bayesian sampler, the blue line pictured represents the analysis 'exploring' the possible values that the parameter might take. The orange line is used to illustrate the concept of multiple 'chains' in an analysis. That is, a Bayesian sampler will generally have multiple processes (chains) exploring the distribution at the same time and will then combine the results of each once finished.



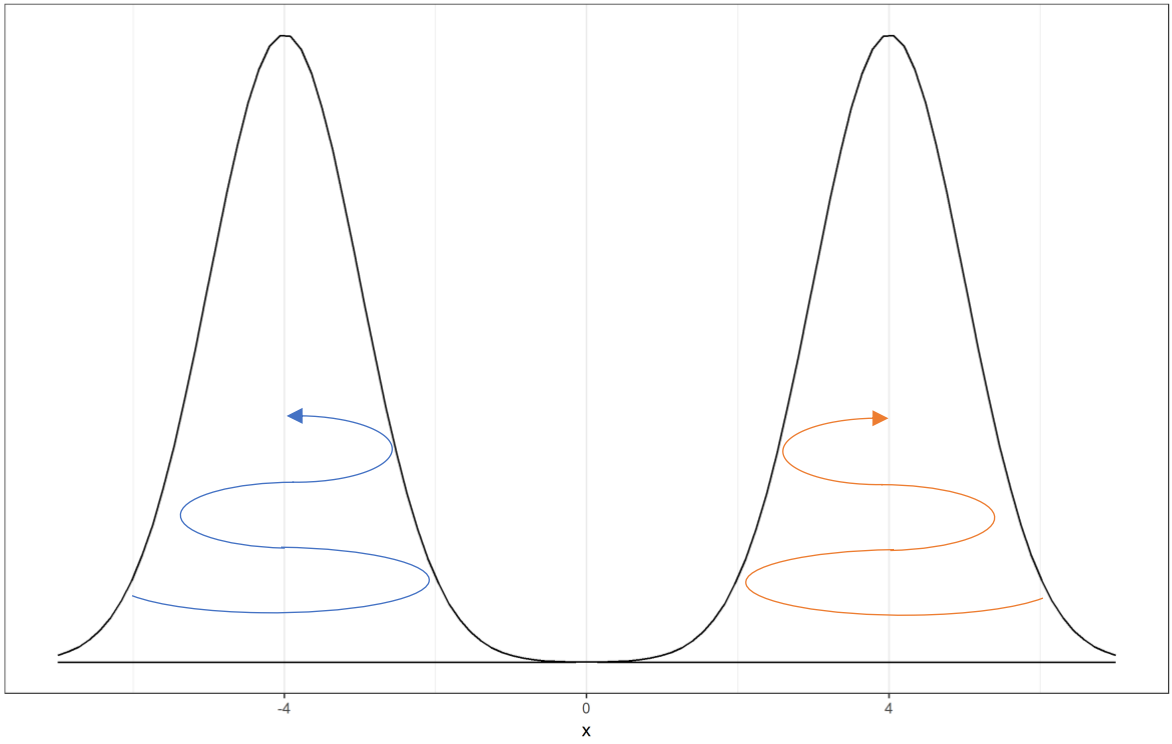
The longer that a given analysis is run, the more comprehensively that the sampler will be able to explore the distribution, and the more accurately the results will represent the 'true' (population) parameter values. This 'exploration' however, can also be drawback for a Bayesian samplers if the distribution is not easy to explore.

9.2.2 Reflection Invariance

In a factor analytic context, this drawback first manifests as a consequence of the difficulties with reflection invariance. Reflection invariance is primarily concerned with the 'sign' of the factor loadings in the model (i.e. whether they are positive or negative). Because the likelihood function for a factor model involves squaring the factor loadings, the likelihood is then the same whether the factor loadings are positive or negative. This means that while the magnitude of factor loadings will affect fit to the data, whether those loadings are positive or negative will not. This is relatively easily addressed in the maximum-likelihood framework through the selection of appropriate initial values to optimise. In a Bayesian framework however, this means that there are two distributions to explore for each factor

loading: one positive and one negative. This is where the term ‘reflection’ invariance comes from, as the values are the same, simply reflected about 0.

This becomes especially problematic for Bayesian analyses with multiple chains, as depicted below, where one chain explores the positive values for the loadings and one explores the negative values:



If one chain has explored the positive values of the factor loadings, and one the negative, combining their results at the end of an analysis will provide severely biased results (especially when estimating means or standard deviations of these distributions).

A commonly recommended approach is to constrain some of the factor loadings in the model to only take on positive values (or only negative). However, this needs to be done on some substantive/theoretical basis, as this alters the interpretation of the final model. The factor loadings in the current thesis are estimating the relationship between motor symptom severity and severity ratings on the MDS-UPDRS. In this case, it is substantively reasonable

to assume that (for some items only) higher levels of symptom severity will result in a higher severity rating on the MDS-UPDRS. That is, there is substantive justification to constrain some loadings to be positive. There is not, however, justification to constrain all loadings to be positive, as some motor symptoms may actually result in other symptoms being rated as less severe. For example, individuals with higher levels of rigidity may exhibit lower levels of resting tremor. As such, it will be sufficient to only constrain the loadings for MDS-UPDRS items on the motor symptoms that they are specifically intended to assess. As an example, for the MDS-UPDRS items that assess rigidity (3.3a-3.3e), only their loadings on the latent rigidity factor will be constrained to positivity. These partial constraints will be sufficient to restrict the remaining loadings from reflecting about 0.

9.2.3 Sizes and Shapes

Even when there is only one distribution (per parameter) to explore, in some cases these distributions can still be difficult for a Bayesian sampler to explore efficiently. This is best illustrated through emphasising the iterative nature of these samplers. While the previous graphics depicted the exploration of a distribution as a single continuous line, the reality is actually a series of ‘steps’. At each iteration of the analysis, the sampler ‘steps’ to a new part of the distribution. However, for example, if that distribution is very large taking the right size of ‘step’ becomes much more difficult. Taking smaller steps means that the analysis needs to be run for far longer to explore the full distribution. On the other hand, taking larger steps increases the risk that the sampler will step ‘outside’ the distribution and return an invalid value.

The computational effort to find this ‘goldilocks’ step size reduces the efficiency of the analysis, requiring it to be run for longer to return accurate results. In comparison, the standard normal distribution (with mean 0, variance 1) is markedly easier to sample from. To

take advantage of this, a commonly-recommended approach is the use of a non-centered parameterisation.

Under the non-centered parameterisation, we specify the distribution of interest as a function of a standard normal distribution. Thanks to some of the properties of normal distributions, this is relatively simple in most cases. Take for example the normal distribution that we specified as part of our factor analysis:

$$y^* \sim N(\lambda\eta, \varepsilon^2)$$

The mean of a normal distribution (in our case: $\lambda\eta$) is also known as the 'location' parameter. This is because the mean does not influence the shape (or 'width') of a normal distribution, it only influences the position of the distribution on the number line. A useful byproduct of this is that we can reparameterise a normal distribution's mean by adding or subtracting a constant. This means that:

$$y^* \sim N(\lambda\eta, \varepsilon^2)$$

is equivalent to:

$$y^* = \lambda\eta + N(0, \varepsilon^2)$$

We can make a similar change with the variance of the distribution. The variance of a normal distribution is also known as the 'scale' parameter, as it is responsible for determining the range of values covered. Rather than adding or a subtracting a constant, like we did for the mean, we can reparameterise a normal distribution by multiplying by a constant. This means that:

$$y^* \sim N(\lambda\eta, \varepsilon^2)$$

is equivalent to:

$$y^* = N(\lambda\eta, 1) * \varepsilon^2$$

By combining these two properties, we can re-write our distribution in a non-centered parameterisation. That is, we can specify our distribution as a function of the standard normal distribution:

$$y^* \sim N(\lambda\eta, \varepsilon^2)$$

is equivalent to:

$$y^* = \lambda\eta + N(0,1) * \varepsilon^2$$

By using this non-centered parameterisation, the sampler only needs to explore the values of a standard normal distribution (which it can do very efficiently), and those samples are then transformed to the distribution needed for the factor analysis.

9.2.4 Univariate vs. Multivariate

However, this approach is only valid for *univariate* normal distributions, as this reparameterisation is only considering the variance of a given item and not its covariance with other items that had been measured. Take for example the case where there are residual covariances between the observed MDS-UPDRS items. In this case, the distribution of the y^* are then multivariate normal with residual variance-covariance matrix Ψ:

$$y^* \sim MVN(\lambda\eta, \Psi)$$

For the non-centered parameterisation of a multivariate normal distribution, we use the same approach for the mean (i.e. adding as a constant). However, rather than multiplying by the individual variances, we take the Cholesky decomposition of the variance-covariance matrix and multiply it by the vector of standard normal variables:

$$L(\Psi) = cholesky(\Psi)$$

$$y^* = \lambda\eta + L(\Psi) * N(0,1)$$

This provides the same required multivariate normal distribution, without having to directly sample from it (which may be computationally intractable in some cases).

9.3 Appendix C: Specifying Models in Stan

9.3.1 Structure of a Stan Model

The syntax for specifying models in Stan differs from that used in other Bayesian programs in several key ways. Stan separates a given model into distinct components (or 'blocks'), each of which are evaluated in different ways or are meant for different content.

The program blocks of a Stan model are as follows:

```
data {
}

transformed data {
}

parameters {
}

transformed parameters {
}

model {
}

generated quantities {
}
```

For those unfamiliar with Stan, it will likely be more helpful to explain the purpose and uses of each block separately. Appendix D will then illustrate the use of each block with an annotated version of the 6-Factor Lateral Bifactor MDS-UPDRS model from Chapter 3.

9.3.2 Data

```
data {
}
```

The *data* block is where we specify the structure and type of the input data that is going to be passed to the program. We are not able to specify any new data here (i.e., N=248), we can only describe the data that is going to be externally supplied. Stan also does not allow any missingness in the supplied data. This does not mean that the missing values cannot be

accounted for, just that the format of the input data needs to be changed to allow it. This is illustrated in Appendix D.

9.3.3 Transformed Data

```
transformed data {
}
```

The *transformed data* block is where we can specify new data to be used by the program. These can be transformations/functions of externally supplied data (e.g., mean-centering covariates). Alternatively, we can also use this block to create new data structures to use in the model rather than having to create them externally and pass them to Stan as data (e.g. an identity matrix).

9.3.4 Parameters

```
parameters {
}
```

The *parameters* block is we specify the parts of the model that will have a distribution to be sampled from. We are not able to directly specify the distributions in this block, only the names and types of parameters that will be sampled.

9.3.5 Transformed Parameters

```
transformed parameters {
}
```

Similar to the transformed data block, *transformed parameters* is where we specify new variables that are functions of the parameters that have been sampled. An example of this which will be illustrated in Appendix D is the non-centered parameterisation for a distribution. This is where we sample a parameter from a standard normal distribution (mean 0, variance 1) and then transform it so that it has the distribution required for our model.

9.3.6 Model

```
model {
}
```

The *model* block is where we specify distributions for those variables that were specified in the parameters block. We are also able to create new variables here (in the same way as the

transformed parameters block), but those variables are regarded as ‘temporary’ and cannot be used by other blocks or saved.

9.3.7 Generated Quantities

```
generated quantities {
}
```

The *generated quantities* block is where we create new variables to be saved after the analysis. We can use this block to generate new observations based on the model, or to create and save information for model fit assessment. While conceptually very similar to the transformed parameters block, generated quantities differs in that the created variables cannot be accessed or used in any other parts of the model. This is because Stan views the generated quantities block as separate from the sampling process. In contrast, the variables created in the transformed parameters are used to inform the sampling process of the overall model.

9.4 Appendix D – Annotated Stan Factor Model

Similar to Appendix C, each block of the model will be presented and explained separately. The model will then be presented in full in Appendix E.

9.4.1 Data

```
data {
  int N;          //Number of individuals
  int J;          //Number of observed items
  int T;          //Number of thresholds
  int E;          //Number of latent factors
  int M;          //Total number of observed responses
  int ii[M];
  int jj[M];
  int y[M];
}
```

This block is where we describe the type and structure of the data that will be supplied to the program. The first half of the block is relatively straightforward, we have:

- N - Our sample size
- J - The number of MDS-UPDRS items in the model
- T - The number of thresholds we need to estimate for each item
- E - The number of latent motor symptom factors

The second half of the block is less intuitive and requires some background in how missing data needs to be prepared for Stan. Because Stan does not allow any missingness in the input data, we need to change the structure of our data so that we are only supplying valid (non-missing) values. However, we also need to track where the missing data is (i.e., which items for which individuals) so that we can account for it in the model. To do this, we need to change our data from a 'wide' format to a 'long' format.

Below is a depiction of how (part of) our dataset would look in 'wide' format (with missingness) and in 'long' format (with no missingness).

Wide Format

		J		
		1	2	3
N	1	y_1		y_2
	2		y_3	y_4
	3	y_5	y_6	

Long Format

N	*J*	*y*
1	1	y_1
1	3	y_2
2	2	y_3
2	3	y_4
3	1	y_5
3	2	y_6

In the ‘wide’ format there is one row for each individual (for a total of *N* rows) and one column for each MDS-UPDRS item (for a total of *J* columns). In the ‘long’ format, however, there is one row *per response*. That is, if an individual was rated on all 33 items of the MDS-UPDRS motor assessment, they would have 33 rows. If instead they were only rated on 20 items, then they would have 20 rows of observations. For each row, the variable *N* indicates which individual the rating belongs to, the variable *J* indicates which item the rating is on, and the variable *y* contains the rating itself.

Returning to our data statement, we can see each of these constructs in the declaration:

- M – Number of responses (i.e. number of rows in ‘long’ format)
- ii**[**M**]** – A column of integers (with M rows) identifying which individual each rating belongs to
- jj**[**M**]** – A column of integers (with M rows) identifying which MDS-UPDRS item the rating was on
- y**[**M**]** – A column of integers (with M rows) containing the MDS-UPDRS ratings

9.4.2 Transformed Data

```
transformed data{
//Factor Loadings
  int right[11]  = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11]   = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8]   = {1,2,18,19,20,21,22,23};
  int rtrem[6]   = {28,29,30,31,32,33};
  int rigid[5]   = {3,4,5,6,7};
  int akin[6]    = {8,9,10,11,12,13};
  int lakin[4]   = {14,15,16,17};
  int ktrem[4]   = {24,25,26,27};
//Cross Loadings
  int r_n[22]    = {1,2,3,5,7,9,11,13,15,17,18,19,
                    20,21,22,23,25,27,29,31,32,33};
  int l_n[22]    = {1,2,3,4,6,8,10,12,14,16,18,19,
                    20,21,22,23,24,26,28,30,32,33};
  int ax_n[25]   = {3,4,5,6,7,8,9,10,11,12,13,14,
                    15,16,17,24,25,26,27,28,29,30,
                    31,32,33};
  int rt_n[27]   = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                    15,16,17,18,19,20,21,22,23,24,
                    25,26,27};
  int rig_n[28]  = {1,2,8,9,10,11,12,13,14,15,16,17,
                    18,19,20,21,22,23,24,25,26,27,
                    28,29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,
                    20,21,22,23,24,25,26,27,28,29,
                    30,31,32,33};
  int lak_n[29]  = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,
                    19,20,21,22,23,24,25,26,27,28,
                    29,30,31,32,33};
  int ktr_n[29]  = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                    15,16,17,18,19,20,21,22,23,28,
                    29,30,31,32,33};
//Prior SDs
  real i_sd = 5.0;
  real u_sd = 0.1;
}
```

In the first section of the transformed data block, we are specifying which MDS-UPDRS items belong to each latent motor symptom factor. We essentially create a ‘list’ of the item numbers for each latent factor and then give that list a name (also known as an ‘array’). We then also create an array of the item numbers for those items that *do not* load on each latent factor (i.e., the cross-loadings). Specifying the item numbers for each factor in this way greatly simplifies the process of constructing the factor loading matrix later in the model. It also allows a great deal of flexibility in specifying which MDS-UPDRS items load onto which motor symptom factors; which further simplifies the process of investigating different factor structures.

The final section of the transformed data block is where we specify the standard deviations for our intended factor loadings and thresholds (‘i_sd’) and for the cross-loadings (‘u_sd’).

9.4.3 Parameters

```
parameters{
  vector[E] eta_raw[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr;
  cholesky_factor_corr[J] u_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[6] akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[27] akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}
```

The first declarations are for the ‘hyperparameters’ in the model; these are the mean and standard deviation for the thresholds (‘thresh_mu’ and ‘thresh_sigma’) and the intended factor loadings (‘load_mu’ and ‘load_sigma’). By estimating the mean and standard deviation for the loadings and thresholds (rather than simply using weakly-informative priors), the model is better able to account for the fact that the thresholds on each item are likely to be similar in some fashion. This helps the sampling process as it provides additional information as to the likely values of each parameter.

Note that the standard deviation hyperparameters are both restricted to be positive, because a negative variance is not statistically possible. This restriction means that the sampler will only explore valid values, and not get ‘stuck’ sampling invalid values.

The next declarations are for the cholesky factors for the correlations between the latent motor symptom factors (‘e_Lcorr’) and between the MDS-UPDRS severity variables (‘u_Lcorr’). There are a number of benefits to sampling the cholesky factor of a correlation matrix, rather than the correlation matrix itself. Firstly, it allows for the use of the multivariate non-centered parameterisation to more efficiently sample from a complex distribution (as covered in Appendix A). Secondly, the sampling of a cholesky factor is more stable than that of the full correlation matrix, because a cholesky factor will always resolve to a symmetric positive-definite matrix whereas sampling the matrix directly will result in invalid samples at some iterations.

Next we have the declarations for the ‘raw’ variables for each individuals’ latent factor scores (‘eta_raw’) and MDS-UPDRS severity scores (‘y_star_raw’). These are denoted as ‘raw’ variables, because they are the samples from the standard normal distribution that will be transformed to the required distributions (i.e. as part of the non-centered parameterisation).

Then we have the declarations for the intended factor loadings (‘right_load’ – ‘ktrem_load’). These loadings are all restricted to be positive, so as to avoid the difficulties with reflection invariance common in Bayesian factor models (as discussed in Appendix B). Finally, we have the declarations for the cross-loadings for each factor (‘right_othl’ – ‘ktrem_othl’). These cross-loadings are not restricted to be positive or negative, as we have no substantive justification to do so. By declaring the cross-loadings and intended loadings as individual vectors, we are able to place unique constraints on each set before constructing the final loading matrix. It also improves the readability of the code itself, as it is much clearer which priors and restrictions are being placed on which symptoms.

9.4.4 Transformed Parameters

```
transformed parameters {
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[right,1] = right_load;
  lam[left,2] = left_load;
  lam[axial,3] = ax_load;
  lam[rtrem,4] = rt_load;
  lam[rigid,5] = rig_load;
  lam[akin,6] = akin_load;
  lam[lakin,7] = lakin_load;
  lam[ktrem,8] = ktrem_load;

  lam[r_n,1] = right_othl;
  lam[l_n,2] = left_othl;
  lam[ax_n,3] = ax_othl;
  lam[rt_n,4] = rt_othl;
  lam[rig_n,5] = rig_othl;
  lam[akin_n,6] = akin_othl;
  lam[lak_n,7] = lakin_othl;
  lam[ktr_n,8] = ktrem_othl;

  lam_ecorr = lam * e_Lcorr;

  for(n in 1:N)
    y_star[n] = lam_ecorr * eta_raw[n] + u_Lcorr * y_star_raw[n];
}
```

The transformed parameters block is where we start 'putting together' the more complex parameters in the model.

The first is the matrix of factor loadings ('lam'). This is where we use the arrays (lists of numbers) that we created in the transformed data block. First, we select the entries in the matrix that correspond to the loadings for each motor symptom factor (e.g., 'lam[right,1]'), and then copy in the loadings that were declared in the parameters block (e.g., '= right_load'). This approach of declaring each 'set' of loadings separately and manually constructing the matrix allows us to specify individual constraints (i.e. the intended loadings are restricted to be positive, but not the cross loadings) and to have different prior distributions (i.e., weakly-informative for the intended loadings and strongly informative for the cross-loadings.

Next, we create a new matrix 'lam_ecorr' which is the product of the loading matrix ('lam') and the cholesky factor for the motor symptom factor correlations ('e_Lcorr'). This is done to essentially 'cache' the calculation so that it can be used later in the construction of

‘y_star’ (the underlying motor symptom severity score). Without this caching, the matrix multiplication would be performed N times in the subsequent loop – an unnecessary computation cost.

Finally, the underlying MDS-UPDRS severity score (‘y_star’) is created from the ‘raw’ variables. This step is actually constructing the latent motor symptom factor (‘eta’) at the same time, but by including it within the construction of ‘y_star’ we save the need to create another variable to hold the result.

9.4.5 Model

```
model{
  e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load ~ normal(load_mu,load_sigma);
  left_load ~ normal(load_mu,load_sigma);
  ax_load ~ normal(load_mu,load_sigma);
  rt_load ~ normal(load_mu,load_sigma);
  rig_load ~ normal(load_mu,load_sigma);
  akin_load ~ normal(load_mu,load_sigma);
  lakin_load ~ normal(load_mu,load_sigma);
  ktrem_load ~ normal(load_mu,load_sigma);

  right_othl ~ normal(0,u_sd);
  left_othl ~ normal(0,u_sd);
  ax_othl ~ normal(0,u_sd);
  rt_othl ~ normal(0,u_sd);
  rig_othl ~ normal(0,u_sd);
  akin_othl ~ normal(0,u_sd);
  lakin_othl ~ normal(0,u_sd);
  ktrem_othl ~ normal(0,u_sd);

  u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N) {
    y_star_raw[n] ~ std_normal();
    eta_raw[n] ~ std_normal();
  }

  thresh_mu ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}
```

Now that we have declared the structure of the different parameters and specified how they are going to be put together, we can declare the relevant prior distributions. First is the prior for the cholesky factor of the correlations between the latent motor symptom factors (‘e_Lcorr’). This cholesky factor is given an LKJ prior with a shape parameter of one, this prior is equivalent to a uniform prior on each individual correlation (i.e., fully uninformative). This prior distribution provides no support for any particular value, considering all possible correlations equally likely. In other words, we have no hypotheses about the strength of relationship between the latent motor symptoms.

Next are the priors for the hyperparameters, these are given weakly-informative priors using the SD specified in the transformed data section. For the means of the loadings and thresholds (‘load_mu’ and ‘thresh_mu’, respectively), normal distributions with a mean of 0 and SD of 5 was used. These priors hypothesise that the mean is 99% likely to fall in the range [-15, 15]. Similarly, for the SD’s of the loadings and thresholds (‘load_sigma’ and ‘thresh_sigma’), a half-cauchy prior was used. While the prior was specified as ‘cauchy’, because the parameters were restricted to being positive, this is equivalent to a half-cauchy.

Next are the prior specifications for the intended factor loadings (‘right_load’ – ‘ktrem_load’). These are using normal distributions with the mean and standard deviation that estimated earlier. Following this, are the priors for the cross-loadings (‘right_othl’ – ‘ktrem_othl’). These are given normal distributed priors with a mean of 0 and a standard deviation of 0.1, hypothesising that 99% of values range between [-0.3, 0.3].

Following this, we specify the prior distribution for the correlations between the MDS-UPDRS items (‘u_Lcorr’). This prior also uses the LKJ distribution (like that used for the correlations between the latent motor symptom factors), but with a shape parameter of 10. Increasing the shape parameter of the LKJ distribution provides increasingly stronger prior information that each correlation should be 0 (thus allowing the model to be identified).

The next step is to declare the standard normal ‘raw’ variables (‘y_star_raw’ and ‘eta_raw’) that are going to be used as part of the non-centered parameterisation. These variables are then transformed to ‘y_star’ and ‘eta’ in the Transformed Parameters section.

Next, the thresholds for each MDS-UPDRS item are estimated, with the mean and standard deviation (‘thresh_mu’ and ‘thresh_sigma’) that were estimated earlier in the model block.

Finally, the model is tied together with the observed data. The observed MDS-UPDRS items (‘y’) are specified to follow an ordered logistic distribution, with underlying severity score ‘y_star’ and ordered cutpoints ‘thresholds’

9.4.6 Generated Quantities

```
generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                           thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
}
```

The Generated Quantities block is where we save the information needed for estimating model fit statistics during post-processing. The log-likelihood for each observation is saved in the variable ‘log_lik’. This will later be used for estimating the Leave One Out (LOO) fit statistic. Next, the deviance statistic is saved. This will later be used for estimating the Deviance Information Criterion (DIC) fit statistic.

9.5 Appendix E – Study 2 Full Factor Model

```
data{
  int N;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
}

transformed data{
//Factor Loadings
  int right[11]  = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11]   = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8]   = {1,2,18,19,20,21,22,23};
  int rtrem[6]   = {28,29,30,31,32,33};
  int rigid[5]   = {3,4,5,6,7};
  int akin[6]    = {8,9,10,11,12,13};
  int lakin[4]   = {14,15,16,17};
  int ktrem[4]   = {24,25,26,27};
//Cross Loadings
  int r_n[22]    = {1,2,3,5,7,9,11,13,15,17,18,19,20,21,22,23,25,27,29,31,32,33};
  int l_n[22]    = {1,2,3,4,6,8,10,12,14,16,18,19,20,21,22,23,24,26,28,30,32,33};
  int ax_n[25]   = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                    17,24,25,26,27,28,29,30,31,32,33};
  int rt_n[27]   = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                    17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28]  = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,
                    20,21,22,23,24,25,26,27,28,29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,
                    20,21,22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29]  = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,
                    22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29]  = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                    16,17,18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] eta_raw[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr;
  cholesky_factor_corr[J] u_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[6] akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
```

```
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[27] akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters {
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[right,1]  = right_load;
  lam[left,2]   = left_load;
  lam[axial,3]  = ax_load;
  lam[rtrem,4]  = rt_load;
  lam[rigid,5]  = rig_load;
  lam[akin,6]   = akin_load;
  lam[lakin,7]  = lakin_load;
  lam[ktrem,8]  = ktrem_load;

  lam[r_n,1]    = right_othl;
  lam[l_n,2]    = left_othl;
  lam[ax_n,3]   = ax_othl;
  lam[rt_n,4]   = rt_othl;
  lam[rig_n,5]  = rig_othl;
  lam[akin_n,6] = akin_othl;
  lam[lak_n,7]  = lakin_othl;
  lam[ktr_n,8]  = ktrem_othl;

  lam_ecorr = lam * e_Lcorr;

  for(n in 1:N)
    y_star[n] = lam_ecorr * eta_raw[n] + u_Lcorr * y_star_raw[n];
}

model{
  e_Lcorr    ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load ~ normal(load_mu,load_sigma);
  left_load  ~ normal(load_mu,load_sigma);
  ax_load    ~ normal(load_mu,load_sigma);
  rt_load    ~ normal(load_mu,load_sigma);
  rig_load   ~ normal(load_mu,load_sigma);
  akin_load  ~ normal(load_mu,load_sigma);
  lakin_load ~ normal(load_mu,load_sigma);
  ktrem_load ~ normal(load_mu,load_sigma);

  right_othl ~ normal(0,u_sd);
  left_othl  ~ normal(0,u_sd);
  ax_othl    ~ normal(0,u_sd);
  rt_othl    ~ normal(0,u_sd);
  rig_othl   ~ normal(0,u_sd);
  akin_othl  ~ normal(0,u_sd);
  lakin_othl ~ normal(0,u_sd);
  ktrem_othl ~ normal(0,u_sd);

  u_Lcorr    ~ lkj_corr_cholesky(20);

  for(n in 1:N) {
    y_star_raw[n] ~ std_normal();
    eta_raw[n]    ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
```

```
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}

generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] |
y_star[ii[m],jj[m]],thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
}
```

9.6 Appendix F – Study 2 Model Distributions

$$y_{ij} \sim ordered\ logistic\left(y_{ij}^{*}, \tau_j\right)$$

$$\tau_j \sim normal(\alpha, \varepsilon)$$

$$\alpha \sim normal(0, 5)$$

$$\varepsilon \sim half\ cauchy(0{,}5)$$

$$y_i^{*} \sim MVN(\Lambda\eta_i, \Psi)$$

$$\Psi \sim lkj(20)$$

$$\eta_i \sim MVN\left(\vec{0}, \Sigma\right)$$

$$\Sigma \sim lkj(1)$$

$$\lambda_j(cross\ loading) \sim normal(0{,}0.1)$$

$$\lambda_j(intended\ loading) \sim normal(\gamma, \delta)$$

$$\gamma \sim normal(0, 5)$$

$$\delta \sim half\ cauchy(0{,}5)$$

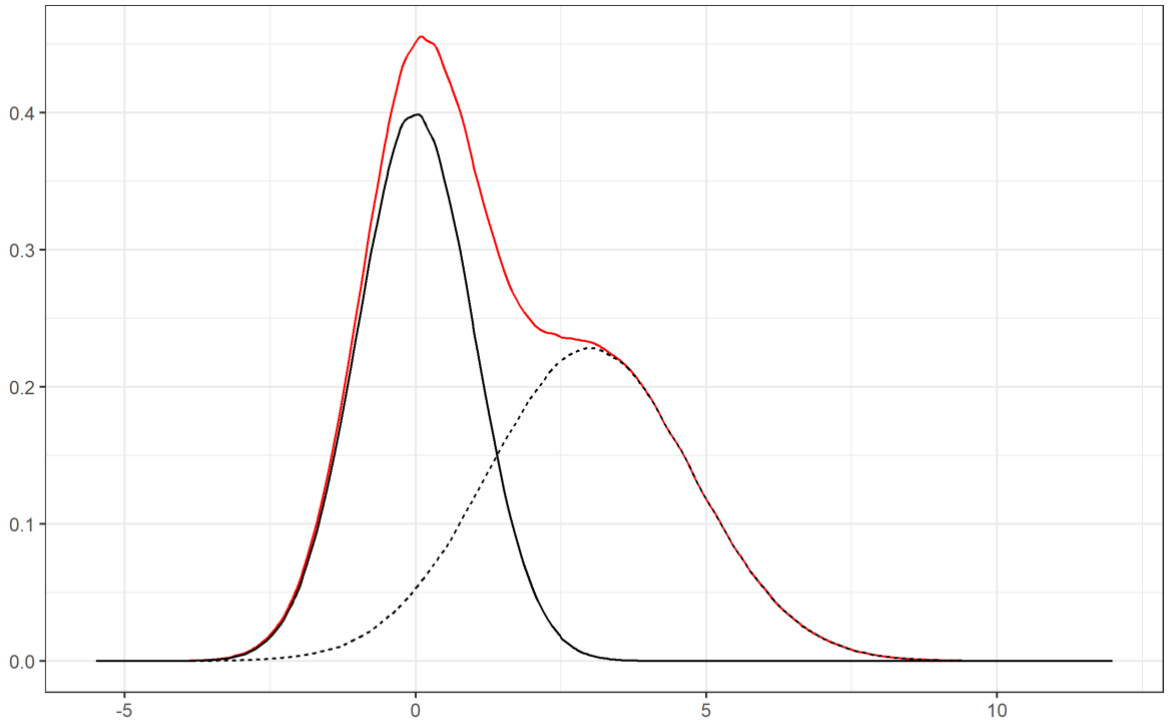
9.7 Appendix G – Parameterising Latent Profile Models

9.7.1 Mixtures of Distributions

The key difference between the models used in Study 2 (MDS-UPDRS measurement) and Study 3 (PD subtyping) is in the distribution of the latent motor symptoms, the variable η for each individual. In a standard factor model (such as the one used in Study 2) we specify η as arising from a single multivariate normal distribution with mean μ and variance-covariance matrix $\sum$:

$$\eta \sim MVN(\mu, \Sigma)$$

A latent profile analysis, on the other hand, supposes that heterogeneity in the data is due to the presence of *multiple* distributions underlying these latent factors. This depicted in the figure below, where the observed non-normality is due to the presence of two underlying distributions:



To parameterise a model like this we need to account for both the mean and covariance structures for each distribution, as well as the probability of belonging to each of the distributions in the sample.

For each of the *K* distributions in the sample, we have the probability of belonging to that subtype θ_k, the vector of means specific to that subtype μ_k, and the variance-covariance matrix specific to that subtype Σ_k. This means that there will be *K* multivariate normal distributions for each individual.

The distribution of η is then the sum of each distribution weighted by its probability:

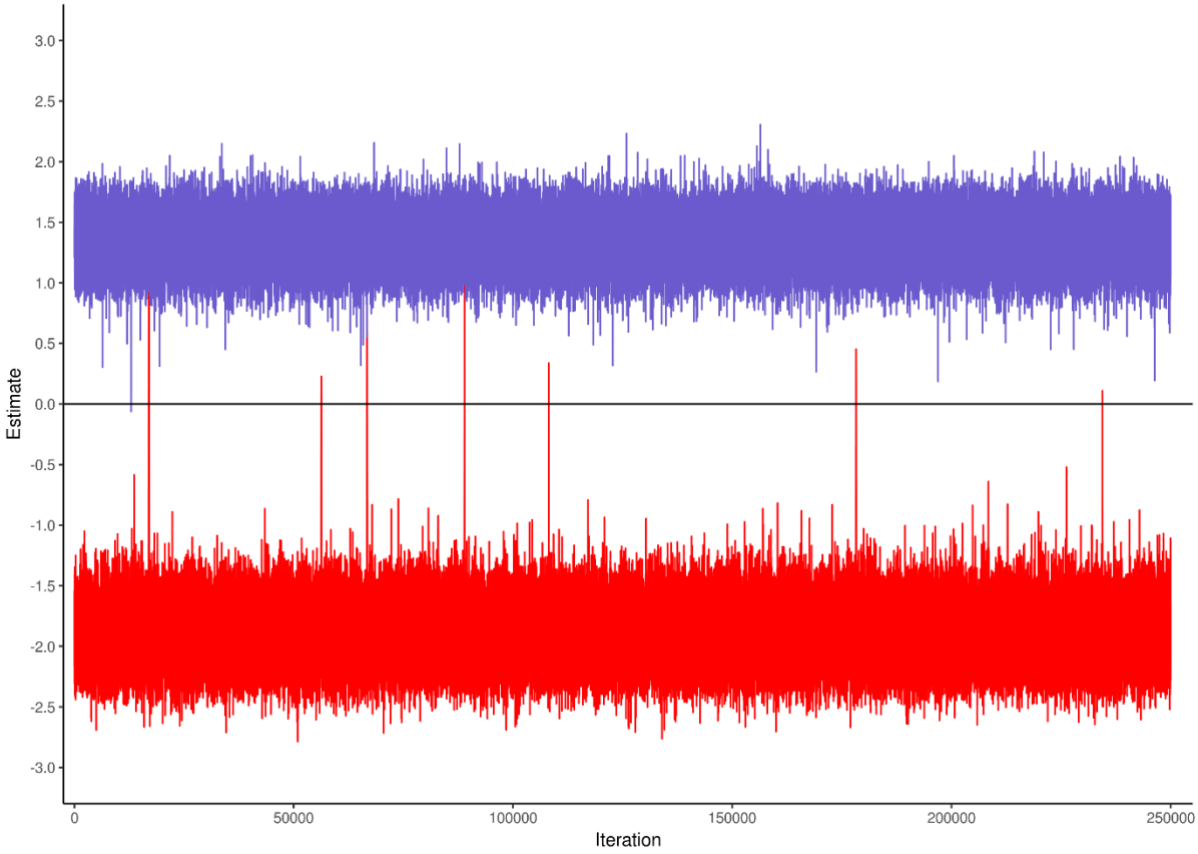
$$\sum \theta_k MVN(\mu_k, \Sigma_k)$$

9.8 Appendix H – Difficulties Estimating Mixture Models

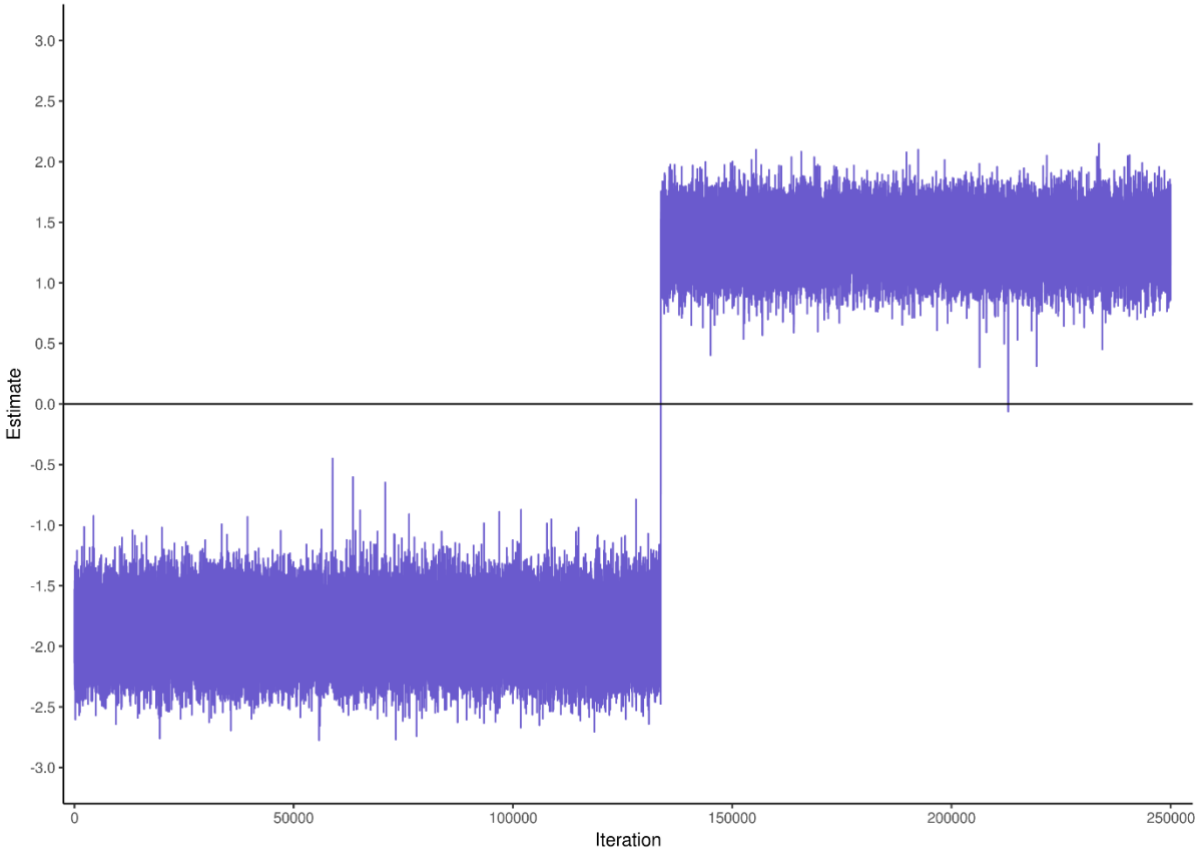
9.8.1 Label Switching

The primary difficulty associated with the Bayesian estimation of mixture models (such as LPA) is that of label switching. The appearance of label switching can be very similar to that of reflection invariance discussed in Appendix B, but the underlying causes are quite different. Label switching is caused by the presence of multiple distributions underlying a given parameter (η in the current analysis). While multiple distributions may be present, there is no meaningful 'ordering' to them. What the analysis labels as the 'first' distribution, 'second' distribution, etc., can change from one iteration to the next (hence the name 'label switching').

There are two forms of label-switching: between-chain and within-chain. Recalling that a Bayesian analysis often involves multiple 'chains' of estimates, between-chain label switching occurs when different chains label different distributions as the 'first'. This is best illustrated graphically. Take for example a mixture model of simulated data with two distributions, where we know that the first has a mean of 1.5, and the second has a mean of -2. If we then conduct an analysis with two chains, and inspect the estimated mean of the first distribution, we would expect both chains to have estimated the mean as 1.5. Upon inspecting the below traceplot of the estimated values of the first distribution's mean, we can see that one chain has estimated the first distribution's mean as 1.5, and the other chain has estimated the mean as -2. In other words, the two chains have labelled different distributions as the 'first' distribution in the sample.



In contrast, within-chain label-switching occurs when a single chain changes its distribution labelling. This is depicted in the traceplot below, where a single chain is estimating the mean of the second distribution (which we simulated to be -2.0). The traceplot shows that the first half of the estimations have correctly labelled the second distribution and estimated the correct mean. However, the second half of estimates have identified the wrong distribution as the second distribution in the dataset, and so there is a distinct ‘switch’ from one distribution to another in the dataset.



When any kind of label-switching occurs, the common practice is to impose an artificial ordering on the labelling of distributions. The standard approach is to order the labelling by the size of the distribution means (i.e., the distribution with the smallest mean will be labelled as the 'first' distribution, and so on). However, label-switching can still occur with this type of labelling. If, for example, two distributions have means that are close together in size, one distribution's mean may not always remain larger than the other's during the sampling process and so the labels will 'switch' to follow the imposed ordering. When this occurs, further constraints can be imposed to correct the labelling.

For the subtyping analyses in the present thesis, models were estimated using a single chain to avoid between-chain label switching and, if within-chain label switching occurred, the subtypes were ordered by their means.

9.9 Appendix I – Ch.4 Annotated Stan Factor Model

Similar to Appendix D, each block of the model will be presented and explained separately. The model will then be presented in full in Appendix K.

9.9.1 Data

```
data{
  int N;
  int J;
  int T;
  int E;
  int K;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
  vector[N] dur;
  vector[N] led;
}
```

This block is almost identical to that of the BSEM model in Ch. 3, with two key differences. First, there is now an integer ‘K’ to represent the number of latent classes (subtypes) to estimate. By passing the number of latent classes as data, rather than ‘hard-coding’ it into the model we can use the same base syntax for all models, and only change the data being passed to the model. Secondly, we also now have vectors containing the disease duration and levodopa equivalent dose for each individual. These responses are standardised before being passed to the model, as this is easier for the Hamiltonian Monte Carlo algorithm.

9.9.2 Transformed Data

```
transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int akin[6] = {8,9,10,11,12,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,
                 21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,
                 21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,
                  16,17,24,25,26,27,28,29,30,31,32,
                  33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                  15,16,17,18,19,20,21,22,23,24,25,
                  26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,
                   18,19,20,21,22,23,24,25,26,27,28,
                   29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,
                    20,21,22,23,24,25,26,27,28,29,30,
                    31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,
                   19,20,21,22,23,24,25,26,27,28,29,
                   30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                   15,16,17,18,19,20,21,22,23,28,29,
                   30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
  vector[K] prop = rep_vector(10,K);
}
```

This block is also near-identical to that of the BSEM in Ch.3. The only difference here is the addition of a vector ‘prop’. This vector is used to parameterise the Dirichlet prior distribution for the subtype proportions. We use the ‘rep_vector’ function here to create a vector that has the same length as the number of latent classes (as set by the integer ‘K’ passed as data), with a value of 10 in each entry. This allows us to set the Dirichlet prior of $D(10_1,\ldots,10_k)$ for each number of latent classes being estimated.

9.9.3 Parameters

```
parameters{
  row_vector[E] led_beta;
  simplex[K] theta;
  vector[E] eta_mu[K-1];
  vector[E] eta[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr[K];
  cholesky_factor_corr[J] u_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[6] akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[27] akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}
```

The differences in this block stem from the handling of the latent motor symptom factors (eta). Firstly, we have the row vectors (‘led_beta’). These will hold the regression coefficients for the effects of each covariate on each of the motor symptom factors. By declaring these as row vectors (as opposed to vectors), the multiplication of these with the vectors of observed covariates (passed as data) will result in a matrix, which will simplify later operations in the model.

Next, we have the simplex ‘theta’ which contains the proportions for each latent class. A simplex is a vector that is constrained to sum to 1, which is perfect for use as a vector of proportions. Following this is ‘eta_mu’ the array of latent motor symptom means for each subtype. The array is of length ‘K-1’, in other words, one less than the number of latent classes in the sample. This is because the last subtype will have their means fixed to 0 and used as the reference class. Following this, the Cholesky factors for the correlations between

the latent motor symptoms in each class are declared ‘e_Lcorr’. The difference between this declaration and the one in the BSEM model, is the declaration of ‘K’ Cholesky factors, because there needs to be one for each of the K latent classes.

The last change is the declaration of the latent motor symptoms for each individual (‘eta’). In the BSEM model, we used the non-centered parameterisation and so these were declared as the ‘raw’ variables, which were later transformed to the required distribution. When using a mixture of distributions, however, the non-centered parameterisation is not available. As such, ‘eta’ is declared and directly estimated.

9.9.4 Transformed Parameters

```
transformed parameters{
  matrix[J,E] lam;
  vector[E] e_mu[K];
  vector[K] contrib[N];
  vector[J] y_star[N];
  matrix[N,E] linpred = led*led_beta;

  lam[right,1]  = right_load;
  lam[left,2]   = left_load;
  lam[axial,3]  = ax_load;
  lam[rtrem,4]  = rt_load;
  lam[rigid,5]  = rig_load;
  lam[akin,6]   = akin_load;
  lam[lakin,7]  = lakin_load;
  lam[ktrem,8]  = ktrem_load;

  lam[r_n,1]    = right_othl;
  lam[l_n,2]    = left_othl;
  lam[ax_n,3]   = ax_othl;
  lam[rt_n,4]   = rt_othl;
  lam[rig_n,5]  = rig_othl;
  lam[akin_n,6] = akin_othl;
  lam[lak_n,7]  = lakin_othl;
  lam[ktr_n,8]  = ktrem_othl;

  for(k in 1:(K-1))
    e_mu[k] = eta_mu[k];

  e_mu[K] = rep_vector(0.0,E);

  for(n in 1:N)
    for(k in 1:K)
      contrib[n][k]
        = multi_normal_cholesky_lpdf(eta[n] | e_mu[k] + linpred[n]',
                                             e_Lcorr[k]);

  for(n in 1:N)
    y_star[n] = lam * eta[n] + u_Lcorr * y_star_raw[n];
}
```

Firstly, we have the creation of the matrix ‘linpred’, which holds the covariates and associated regression coefficients for each individual. This matrix will then be used in the creation of the mixture distributions for the latent motor symptoms. Next is the creation of the factor loadings matrix ‘lam’, which is unchanged from the BSEM in Ch. 3.

The next step is the creation of different components for the mixture model. We firstly create a new array of vectors to hold the estimated latent means for each class (‘e_mu’), and then set the means of the last class to 0. After this is the array of vectors ‘contrib’. This array will hold the different distributions for each latent class for each individual and will be combined with the latent class proportions ‘theta’ in the model block.

Finally, the non-centered parameterisation for y* is applied.

9.9.5 Model

```
model{
  led_beta ~ normal(0,5);

  for(k in 1:(K-1))
    eta_mu[k] ~ normal(0,5);

  for(k in 1:K)
    e_Lcorr[k] ~ lkj_corr_cholesky(20);

  theta ~ dirichlet(prop);

  target += log_mix(theta,contrib);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load ~ normal(load_mu,load_sigma);
  left_load  ~ normal(load_mu,load_sigma);
  ax_load    ~ normal(load_mu,load_sigma);
  rt_load    ~ normal(load_mu,load_sigma);
  rig_load   ~ normal(load_mu,load_sigma);
  akin_load  ~ normal(load_mu,load_sigma);
  lakin_load ~ normal(load_mu,load_sigma);
  ktrem_load ~ normal(load_mu,load_sigma);

  right_othl ~ normal(0,u_sd);
  left_othl  ~ normal(0,u_sd);
  ax_othl    ~ normal(0,u_sd);
  rt_othl    ~ normal(0,u_sd);
  rig_othl   ~ normal(0,u_sd);
  akin_othl  ~ normal(0,u_sd);
  lakin_othl ~ normal(0,u_sd);
  ktrem_othl ~ normal(0,u_sd);

  u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N)
    y_star_raw[n] ~ std_normal();

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}
```

There are again several similarities between this model and the BSEM of Ch. 3. All of the changes are at the ‘higher’ levels of the model, as it is only the distribution of the latent factors ‘eta’ that is different. We first need to specify the weakly-informative priors for the covariate regression coefficients and latent motor symptom means. Next are the strongly-informative priors on the Cholesky factors for the correlations within each of the subtypes. It is these strongly-informative priors that are allowing us to circumvent the assumption of local independence and still have an identified model. The next step is to specify the weakly-

informative Dirichlet prior distribution for the latent class proportions (using the vector ‘prop’ that was declared in the transformed data block’). Finally, the simplex of latent class proportions ‘theta’ and the array of distributions for each individual ‘contrib’ are combined using the ‘log_mix’ function to form the mixture distribution for ‘eta’.

The rest of the model is the same as that declared for Ch. 3.

9.9.6 Generated Quantities

```
generated quantities {
  vector[K] pt = log(theta);
  vector[K] log_Pr[N];
  matrix[N,K] rel_ent_num;
  real rel_ent;
  vector[M] log_lik;
  real deviance;
  matrix[E,E] e_Corr[K];
  matrix[J,J] u_Corr = multiply_lower_tri_self_transpose(u_Lcorr);

  for(n in 1:N) {
    for(k in 1:K) {
      log_Pr[n,k] = (pt[k] + contrib[n,k]) - log_sum_exp(contrib[n]+pt);
      rel_ent_num[n,k] = log_Pr[n,k] * exp(log_Pr[n,k]);
    }
  }

  rel_ent = 1.0 - (-sum(rel_ent_num))/(N*log(K));

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                           thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);

  for(k in 1:K)
    e_Corr[k] = multiply_lower_tri_self_transpose(e_Lcorr[k]);
}
```

There is slightly more post-processing required for this model than was required for the BSEM of Ch. 3. We first need to estimate the individual log-probabilities of belonging to each of the subtypes in the sample (‘log_Pr’ to be used in the path analysis of Ch. 5). We also need to estimate the Entropy fit statistic (‘rel_ent’) for the quality of classification in the subtyping model. Finally, we need to transform the Cholesky factors in each subtype to a full correlation matrix (‘e_Corr’), so that we can interpret the corellations between each motor symptom in each subtype.

The deviance, log-likelihood, and MDS-UPDRS item correlation matrix are all declared and estimated in the same way as in the Ch.3 BSEM.

9.10 Appendix J – Study 3 Full Model

```
data{
  int N;
  int J;
  int T;
  int E;
  int K;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
  vector[N] dur;
  vector[N] led;
}

transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int akin[6] = {8,9,10,11,12,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,
                 21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,
                 21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,
                  16,17,24,25,26,27,28,29,30,31,32,
                  33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                  15,16,17,18,19,20,21,22,23,24,25,
                  26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,
                   18,19,20,21,22,23,24,25,26,27,28,
                   29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,
                    20,21,22,23,24,25,26,27,28,29,30,
                    31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,
                   19,20,21,22,23,24,25,26,27,28,29,
                   30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                   15,16,17,18,19,20,21,22,23,28,29,
                   30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
  vector[K] prop = rep_vector(10,K);
}

parameters{
  row_vector[E] led_beta;
  simplex[K] theta;
  vector[E] eta_mu[K-1];
  vector[E] eta[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr[K];
  cholesky_factor_corr[J] u_Lcorr;
```

```
  vector[J] y_star_raw[N];

  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[6] akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[27] akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  vector[E] e_mu[K];
  vector[K] contrib[N];
  vector[J] y_star[N];
  matrix[N,E] linpred = led*led_beta;

  lam[right,1]  = right_load;
  lam[left,2]   = left_load;
  lam[axial,3]  = ax_load;
  lam[rtrem,4]  = rt_load;
  lam[rigid,5]  = rig_load;
  lam[akin,6]   = akin_load;
  lam[lakin,7]  = lakin_load;
  lam[ktrem,8]  = ktrem_load;

  lam[r_n,1]    = right_othl;
  lam[l_n,2]    = left_othl;
  lam[ax_n,3]   = ax_othl;
  lam[rt_n,4]   = rt_othl;
  lam[rig_n,5]  = rig_othl;
  lam[akin_n,6] = akin_othl;
  lam[lak_n,7]  = lakin_othl;
  lam[ktr_n,8]  = ktrem_othl;

  for(k in 1:(K-1))
    e_mu[k] = eta_mu[k];

  e_mu[K] = rep_vector(0.0,E);

  for(n in 1:N)
    for(k in 1:K)
      contrib[n][k]
        = multi_normal_cholesky_lpdf(eta[n] | e_mu[k] + linpred[n]',
                                             e_Lcorr[k]);

  for(n in 1:N)
    y_star[n] = lam * eta[n] + u_Lcorr * y_star_raw[n];
}

model{
  led_beta ~ normal(0,5);

  for(k in 1:(K-1))
    eta_mu[k] ~ normal(0,5);

  for(k in 1:K)
```

```
    e_Lcorr[k] ~ lkj_corr_cholesky(20);

  theta ~ dirichlet(prop);

  target += log_mix(theta,contrib);

  load_mu ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load ~ normal(load_mu,load_sigma);
  left_load  ~ normal(load_mu,load_sigma);
  ax_load    ~ normal(load_mu,load_sigma);
  rt_load    ~ normal(load_mu,load_sigma);
  rig_load   ~ normal(load_mu,load_sigma);
  akin_load  ~ normal(load_mu,load_sigma);
  lakin_load ~ normal(load_mu,load_sigma);
  ktrem_load ~ normal(load_mu,load_sigma);

  right_othl ~ normal(0,u_sd);
  left_othl  ~ normal(0,u_sd);
  ax_othl    ~ normal(0,u_sd);
  rt_othl    ~ normal(0,u_sd);
  rig_othl   ~ normal(0,u_sd);
  akin_othl  ~ normal(0,u_sd);
  lakin_othl ~ normal(0,u_sd);
  ktrem_othl ~ normal(0,u_sd);

  u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N)
    y_star_raw[n] ~ std_normal();

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],[jj[m]]);
}

generated quantities {
  vector[K] pt = log(theta);
  vector[K] log_Pr[N];
  matrix[N,K] rel_ent_num;
  real rel_ent;
  vector[M] log_lik;
  real deviance;
  matrix[E,E] e_Corr[K];
  matrix[J,J] u_Corr = multiply_lower_tri_self_transpose(u_Lcorr);

  for(n in 1:N) {
    for(k in 1:K) {
      log_Pr[n,k] = (pt[k] + contrib[n,k])
        - log_sum_exp(contrib[n]+pt);
      rel_ent_num[n,k] = log_Pr[n,k] * exp(log_Pr[n,k]);
    }
  }

  rel_ent = 1.0 - (-sum(rel_ent_num))/(N*log(K));

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                             thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
```

```
  for(k in 1:K)
    e_Corr[k] = multiply_lower_tri_self_transpose(e_Lcorr[k]);
}
```

9.11 Appendix K – Ch. 4 Model Distributions

$$y_{ij} \sim ordered\ logistic\left(y_{ij}^{*}, \tau_j\right)$$

$$\tau_j \sim normal(\alpha, \varepsilon)$$

$$\alpha \sim normal(0,5)$$

$$\varepsilon \sim half\ cauchy(0{,}5)$$

$$y_i^{*} \sim MVN(\Lambda\eta_i, \Psi)$$

$$\Psi \sim lkj(20)$$

$$\lambda_j(cross\ loading) \sim normal(0{,}0.1)$$

$$\lambda_j(intended\ loading) \sim normal(\gamma, \delta)$$

$$\gamma \sim normal(0,5)$$

$$\delta \sim half\ cauchy(0{,}5)$$

$$\eta_i \sim \sum \theta_k MVN(\mu_k, \Sigma_k)$$

$$\theta \sim D(10_1, \ldots, 10_k)$$

$$\mu_{k<K} \sim normal(0,5)$$

$$\mu_K = \vec{0}$$

$$\Sigma_k \sim lkj(20)$$

10 Technical Appendices

10.1 Technical Appendix A – Dirichlet Log-Probability Function

```
#ifndef STAN_MATH_PRIM_MAT_PROB_DIRICHLET_LPMF_HPP
#define STAN_MATH_PRIM_MAT_PROB_DIRICHLET_LPMF_HPP

#include <stan/math/prim/scal/meta/include_summand.hpp>
#include <stan/math/prim/scal/meta/partials_return_type.hpp>
#include <stan/math/prim/scal/meta/return_type.hpp>
#include <stan/math/prim/scal/err/check_consistent_sizes.hpp>
#include <stan/math/prim/scal/err/check_positive.hpp>
#include <stan/math/prim/mat/err/check_simplex.hpp>
#include <stan/math/prim/mat/fun/lgamma.hpp>
#include <stan/math/prim/mat/fun/digamma.hpp>
#include <stan/math/prim/mat/fun/value_of.hpp>
#include <stan/math/prim/mat/meta/is_vector.hpp>
#include <stan/math/prim/mat/meta/operands_and_partials.hpp>
#include <stan/math/prim/mat/meta/is_constant_struct.hpp>
#include <stan/math/prim/mat/meta/get.hpp>
#include <stan/math/prim/mat/meta/length.hpp>
#include <stan/math/prim/mat/meta/vector_seq_view.hpp>

namespace stan {
namespace math {

/**
 * The log of the Dirichlet density for the given theta and
 * a vector of prior sample sizes, alpha.
 * Each element of alpha must be greater than 0.
 * Each element of theta must be greater than or 0.
 * Theta sums to 1.
 *
 * \f[
 * \theta\sim\mbox{Dirichlet}(\alpha_1,\ldots,\alpha_k)\\
 * \log(p(\theta\,|\,\alpha_1,\ldots,\alpha_k))=\log\left(
 * \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)+
 * \cdots+\Gamma(\alpha_k)}*
 * \left(\theta_1^{\alpha_1-1}+\cdots+\theta_k^{\alpha_k-1}
 * \right)\right)\\
 * =\log(\Gamma(\alpha_1+\cdots+\alpha_k))-\left(\log(
 * \Gamma(\alpha_1))+\cdots+\log(\Gamma(\alpha_k))\right)+
 * (\alpha_1-1)\log(\theta_1)+\cdots+(\alpha_k-1)\log(\theta_k)
 * \f]
 *
 * \f[
 * \frac{\partial }{\partial
 * \theta_x}\log(p(\theta\,|\,\alpha_1,\ldots,\alpha_k))=\frac{\alpha_x-
1}{\theta_x}
 * \f]
 *
 * \f[
 *
\frac{\partial}{\partial\alpha_x}\log(p(\theta\,|\,\alpha_1,\ldots,\alpha_k))=\psi_
{(0)}(\sum\alpha)-\psi_{(0)}(\alpha_x)+\log\theta_x
 * \f]
 *
 * @param theta A scalar vector.
 * @param alpha Prior sample sizes.
 * @return The log of the Dirichlet density.
 * @throw std::domain_error if any element of alpha is less than
 * or equal to 0.
 * @throw std::domain_error if any element of theta is less than 0.
 * @throw std::domain_error if the sum of theta is not 1.
 * @tparam T_prob Type of scalar.
 * @tparam T_prior_size Type of prior sample sizes.
 */
```

```
template <bool propto, typename T_prob, typename T_prior_size>
typename return_type<T_prob, T_prior_size>::type dirichlet_lpmf(
    const T_prob& theta, const T_prior_size& alpha) {
  static const char* function = "dirichlet_lpmf";

  typedef typename stan::partials_return_type<T_prob, T_prior_size>::type
      T_partials_return;
  typedef typename Eigen::Matrix<T_partials_return, -1, 1> T_partials_vec;

  check_consistent_sizes(function, "probabilities", theta, "prior sample
sizes", alpha);
  check_positive(function, "prior sample sizes", alpha);
  check_simplex(function, "probabilities", theta);

  vector_seq_view<T_prob> theta_vec(theta);
  T_partials_vec theta_dbl = value_of(theta_vec[0]);

  vector_seq_view<T_prior_size> alpha_vec(alpha);
  T_partials_vec alpha_dbl = value_of(alpha_vec[0]);

  T_partials_return lp(0.0);

  if (include_summand<propto, T_prior_size>::value)
    lp += lgamma(alpha_dbl.sum()) - lgamma(alpha_dbl).sum();

  if (include_summand<propto, T_prob, T_prior_size>::value)
    lp += (theta_dbl.array().log() * (alpha_dbl.array() - 1.0)).sum();

  T_partials_vec theta_deriv = (alpha_dbl.array() - 1.0) /
                               theta_dbl.array();

  T_partials_vec alpha_deriv = digamma(alpha_dbl.sum())
                             - digamma(alpha_dbl).array()
                             + theta_dbl.array().log();

  operands_and_partials<T_prob, T_prior_size> ops_partials(theta, alpha);
  if (!is_constant_struct<T_prob>::value)
    ops_partials.edge1_.partials_ = theta_deriv;

  if (!is_constant_struct<T_prior_size>::value)
    ops_partials.edge2_.partials_ = alpha_deriv;

  return ops_partials.build(lp);
}

template <typename T_prob, typename T_prior_size>
typename return_type<T_prob, T_prior_size>::type dirichlet_lpmf(
    const T_prob& theta, const T_prior_size& alpha) {
  return dirichlet_lpmf<false>(theta, alpha);
}

}  // namespace math
}  // namespace stan
#endif
```

10.2 Technical Appendix B – Inverse Logit

```
#ifndef STAN_MATH_PRIM_SCAL_FUN_INV_LOGIT_HPP
#define STAN_MATH_PRIM_SCAL_FUN_INV_LOGIT_HPP

#include <stan/math/prim/scal/fun/constants.hpp>
#include <cmath>

namespace stan {
namespace math {

/**
 * Returns the inverse logit function applied to the argument.
 *
 * The inverse logit function is defined by
 *
 * \f$\mbox{logit}^{-1}(x) = \frac{1}{1 + \exp(-x)}\f$.
 *
 * This function can be used to implement the inverse link function
 * for logistic regression.
 *
 * The inverse to this function is <code>logit</code>.
 *
 *
 \f[
 \mbox{inv\_logit}(y) =
 \begin{cases}
 \mbox{logit}^{-1}(y) & \mbox{if } -\infty\leq y \leq \infty \\[6pt]
 \textrm{NaN} & \mbox{if } y = \textrm{NaN}
 \end{cases}
 \f]

 \f[
 \frac{\partial\, \mbox{inv\_logit}(y)}{\partial y} =
 \begin{cases}
 \frac{\partial\, \mbox{logit}^{-1}(y)}{\partial y} & \mbox{if } -\infty\leq
 y\leq \infty \\[6pt] \textrm{NaN} & \mbox{if } y = \textrm{NaN} \end{cases} \f]

 \f[
 \mbox{logit}^{-1}(y) = \frac{1}{1 + \exp(-y)}
 \f]

 \f[
 \frac{\partial \, \mbox{logit}^{-1}(y)}{\partial y} =
 \frac{\exp(y)}{(\exp(y)+1)^2} \f]
 *
 * @param a Argument.
 * @return Inverse logit of argument.
 */
inline double inv_logit(double a) {
  using std::exp;
  if (a < 0) {
    double exp_a = exp(a);
    if (a < LOG_EPSILON)
      return exp_a;
    return exp_a / (1 + exp_a);
  }
  return 1 / (1 + exp(-a));
}

}  // namespace math
}  // namespace stan
#endif
```

10.3 Technical Appendix C – Log-Difference of Two Inverse Logits

10.3.1 Primitive Types

```
#ifndef STAN_MATH_PRIM_SCAL_FUN_LOG_INV_LOGIT_DIFF_HPP
#define STAN_MATH_PRIM_SCAL_FUN_LOG_INV_LOGIT_DIFF_HPP

#include <stan/math/prim/scal/fun/log1m_exp.hpp>
#include <stan/math/prim/scal/fun/log1p_exp.hpp>
#include <boost/math/tools/promotion.hpp>

namespace stan {
namespace math {

/**
 * Returns the natural logarithm of the difference of the
 * inverse logits of the specified arguments.
 *
   \f[
     \mathrm{log\_inv\_logit\_diff}(x,y) =
      \ln\left(\frac{1}{1+\exp(-x)}-\frac{1}{1+\exp(-y)}\right)
   \f]

   \f[
    \frac{\partial }{\partial x} = -\frac{e^x}{e^y-e^x}-\frac{e^x}{e^x+1}
   \f]

   \f[
    \frac{\partial }{\partial x} = -\frac{e^y}{e^x-e^y}-\frac{e^y}{e^y+1}
   \f]
 *
 * @tparam T1 Type of x argument.
 * @tparam T2 Type of y argument.
 * @param x Argument.
 * @param y Argument.
 * @return Result of log difference of inverse logits of arguments.
 */
template <typename T1, typename T2>
inline typename boost::math::tools::promote_args<T1, T2>::type
log_inv_logit_diff(const T1& x, const T2& y) {
  return x - log1p_exp(x) + log1m_exp(y - x) - log1p_exp(y);
}

}  // namespace math
}  // namespace stan

#endif
```

10.3.2 Reverse-Mode Autodiff Types

```
#ifndef STAN_MATH_REV_SCAL_FUN_LOG_INV_LOGIT_DIFF_HPP
#define STAN_MATH_REV_SCAL_FUN_LOG_INV_LOGIT_DIFF_HPP

#include <stan/math/rev/core.hpp>
#include <stan/math/prim/scal/fun/log_inv_logit_diff.hpp>
#include <stan/math/prim/scal/fun/inv_logit.hpp>
#include <stan/math/prim/scal/fun/inv.hpp>
#include <stan/math/prim/scal/fun/expm1.hpp>

namespace stan {
namespace math {

/**
 * Returns the natural logarithm of the difference of the
 * inverse logits of the specified arguments and its gradients.
 *
   \f[
     \mathrm{log\_inv\_logit\_diff}(x,y) =
      \ln\left(\frac{1}{1+\exp(-x)}-\frac{1}{1+\exp(-y)}\right)
   \f]

   \f[
    \frac{\partial }{\partial x} = -\frac{e^x}{e^y-e^x}-\frac{e^x}{e^x+1}
   \f]

   \f[
    \frac{\partial }{\partial x} = -\frac{e^y}{e^x-e^y}-\frac{e^y}{e^y+1}
   \f]
 *
 * @tparam T1 Type of x argument.
 * @tparam T2 Type of y argument.
 * @param x Argument.
 * @param y Argument.
 * @return Result of log difference of inverse logits of arguments
 *          and gradients.
 */
namespace {
class log_inv_logit_diff_vv_vari : public op_vv_vari {
 public:
  log_inv_logit_diff_vv_vari(vari* avi, vari* bvi)
      : op_vv_vari(log_inv_logit_diff(avi->val_, bvi->val_), avi, bvi) {}
  void chain() {
    avi_->adj_
        -= adj_ * (inv(expm1(bvi_->val_ - avi_->val_)) + inv_logit(avi_-
                                                                   >val_));

    bvi_->adj_
        -= adj_ * (inv(expm1(avi_->val_ - bvi_->val_)) + inv_logit(bvi_-
                                                                   >val_));
  }
};

class log_inv_logit_diff_vd_vari : public op_vd_vari {
 public:
  log_inv_logit_diff_vd_vari(vari* avi, double b)
      : op_vd_vari(log_inv_logit_diff(avi->val_, b), avi, b) {}
  void chain() {
    avi_->adj_ -= adj_ * (inv(expm1(bd_ - avi_->val_)) + inv_logit(avi_-
                                                                   >val_));
  }
};

class log_inv_logit_diff_dv_vari : public op_dv_vari {
 public:
  log_inv_logit_diff_dv_vari(double a, vari* bvi)
```

```
      : op_dv_vari(log_inv_logit_diff(a, bvi->val_), a, bvi) {}
  void chain() {
    bvi_->adj_ -= adj_ * (inv(expm1(ad_ - bvi_->val_)) + inv_logit(bvi_->val_));
  }
};
}  // namespace

inline var log_inv_logit_diff(const var& a, double b) {
  return var(new log_inv_logit_diff_vd_vari(a.vi_, b));
}

inline var log_inv_logit_diff(const var& a, const var& b) {
  return var(new log_inv_logit_diff_vv_vari(a.vi_, b.vi_));
}

inline var log_inv_logit_diff(double a, const var& b) {
  return var(new log_inv_logit_diff_dv_vari(a, b.vi_));
}

}  // namespace math
}  // namespace stan
#endif
```

10.3.3 Forward-Mode Autodiff Types

```
#ifndef STAN_MATH_FWD_SCAL_FUN_LOG_INV_LOGIT_DIFF_HPP
#define STAN_MATH_FWD_SCAL_FUN_LOG_INV_LOGIT_DIFF_HPP

#include <stan/math/fwd/core.hpp>
#include <stan/math/prim/scal/fun/log_inv_logit_diff.hpp>
#include <stan/math/prim/scal/fun/inv_logit.hpp>
#include <stan/math/prim/scal/fun/inv.hpp>
#include <stan/math/prim/scal/fun/expm1.hpp>

namespace stan {
namespace math {
/**
 * Returns fvar with the natural logarithm of the difference of the
 * inverse logits of the specified arguments and its gradients.
 *
   \f[
     \mathrm{log\_inv\_logit\_diff}(x,y) =
      \ln\left(\frac{1}{1+\exp(-x)}-\frac{1}{1+\exp(-y)}\right)
   \f]

   \f[
    \frac{\partial }{\partial x} = -\frac{e^x}{e^y-e^x}-\frac{e^x}{e^x+1}
   \f]

   \f[
    \frac{\partial }{\partial x} = -\frac{e^y}{e^x-e^y}-\frac{e^y}{e^y+1}
   \f]
 *
 * @tparam T1 Type of x argument.
 * @tparam T2 Type of y argument.
 * @param x Argument.
 * @param y Argument.
 * @return Fvar with result of log difference of inverse logits of arguments
 *           and gradients.
 */
template <typename T>
inline fvar<T> log_inv_logit_diff(const fvar<T>& x, const fvar<T>& y) {
  return fvar<T>(
      log_inv_logit_diff(x.val_, y.val_),
      -x.d_ * (inv(expm1(y.val_ - x.val_)) + inv_logit(x.val_))
          - y.d_ * (inv(expm1(x.val_ - y.val_)) + inv_logit(y.val_)));
}

template <typename T>
inline fvar<T> log_inv_logit_diff(const fvar<T>& x, double y) {
  return fvar<T>(log_inv_logit_diff(x.val_, y),
                 -x.d_ * (inv(expm1(y - x.val_)) + inv_logit(x.val_)));
}

template <typename T>
inline fvar<T> log_inv_logit_diff(double x, const fvar<T>& y) {
  return fvar<T>(log_inv_logit_diff(x, y.val_),
                 -y.d_ * (inv(expm1(x - y.val_)) + inv_logit(y.val_)));
}
}  // namespace math
}  // namespace stan
#endif
```

10.4 Technical Appendix D – Log-Mixture of Distributions

```
#ifndef STAN_MATH_PRIM_MAT_FUN_LOG_MIX_HPP
#define STAN_MATH_PRIM_MAT_FUN_LOG_MIX_HPP

#include <stan/math/prim/arr/meta/get.hpp>
#include <stan/math/prim/arr/meta/length.hpp>
#include <stan/math/prim/mat/meta/is_vector.hpp>
#include <stan/math/prim/mat/meta/get.hpp>
#include <stan/math/prim/mat/meta/length.hpp>
#include <stan/math/prim/mat/fun/log_sum_exp.hpp>
#include <stan/math/prim/mat/fun/log.hpp>
#include <stan/math/prim/mat/fun/value_of.hpp>
#include <stan/math/prim/mat/meta/vector_seq_view.hpp>
#include <stan/math/prim/mat/meta/broadcast_array.hpp>
#include <stan/math/prim/mat/meta/operands_and_partials.hpp>
#include <stan/math/prim/scal/err/check_bounded.hpp>
#include <stan/math/prim/scal/err/check_not_nan.hpp>
#include <stan/math/prim/scal/err/check_consistent_sizes.hpp>
#include <stan/math/prim/scal/err/check_finite.hpp>
#include <stan/math/prim/scal/meta/partials_return_type.hpp>
#include <stan/math/prim/scal/meta/operands_and_partials.hpp>
#include <stan/math/prim/scal/meta/scalar_seq_view.hpp>
#include <stan/math/prim/scal/meta/return_type.hpp>
#include <stan/math/prim/scal/meta/is_constant_struct.hpp>
#include <stan/math/prim/scal/fun/value_of.hpp>
#include <vector>

namespace stan {
namespace math {

/**
 * Return the log mixture density with specified mixing proportions
 * and log densities.
 *
 * \f[
 * \frac{\partial }{\partial p_x}
 * \log\left(\exp^{\log\left(p_1\right)+d_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+d_n}\right)
 * =\frac{e^{d_x}}{e^{d_1}p_1+\cdot\cdot\cdot+e^{d_m}p_m}
 * \f]
 *
 * \f[
 * \frac{\partial }{\partial d_x}
 * \log\left(\exp^{\log\left(p_1\right)+d_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+d_n}\right)
 * =\frac{e^{d_x}p_x}{e^{d_1}p_1+\cdot\cdot\cdot+e^{d_m}p_m}
 * \f]
 *
 * @param theta vector of mixing proportions in [0, 1].
 * @param lambda vector of log densities.
 * @return log mixture of densities in specified proportion
 */
template <typename T_theta, typename T_lam>
typename return_type<T_theta, T_lam>::type log_mix(const T_theta& theta,
                                                   const T_lam& lambda) {
  static const char* function = "log_mix";
  typedef typename stan::partials_return_type<T_theta, T_lam>::type
      T_partials_return;

  typedef typename Eigen::Matrix<T_partials_return, -1, 1> T_partials_vec;

  const int N = length(theta);

  check_bounded(function, "theta", theta, 0, 1);
  check_not_nan(function, "lambda", lambda);
  check_not_nan(function, "theta", theta);
```

```
  check_finite(function, "lambda", lambda);
  check_finite(function, "theta", theta);
  check_consistent_sizes(function, "theta", theta, "lambda", lambda);

  scalar_seq_view<T_theta> theta_vec(theta);
  T_partials_vec theta_dbl(N);
  for (int n = 0; n < N; ++n)
    theta_dbl[n] = value_of(theta_vec[n]);

  scalar_seq_view<T_lam> lam_vec(lambda);
  T_partials_vec lam_dbl(N);
  for (int n = 0; n < N; ++n)
    lam_dbl[n] = value_of(lam_vec[n]);

  T_partials_return logp = log_sum_exp((log(theta_dbl) + lam_dbl).eval());

  T_partials_vec theta_deriv(N);
  theta_deriv.array() = (lam_dbl.array() - logp).exp();

  T_partials_vec lam_deriv = theta_deriv.cwiseProduct(theta_dbl);

  operands_and_partials<T_theta, T_lam> ops_partials(theta, lambda);
  if (!is_constant_struct<T_theta>::value) {
    for (int n = 0; n < N; ++n)
      ops_partials.edge1_.partials_[n] = theta_deriv[n];
  }

  if (!is_constant_struct<T_lam>::value) {
    for (int n = 0; n < N; ++n)
      ops_partials.edge2_.partials_[n] = lam_deriv[n];
  }

  return ops_partials.build(logp);
}

/**
 * Return the log mixture density given specified mixing proportions
 * and array of log density vectors.
 *
 * \f[
 * \frac{\partial }{\partial p_x}\left[
 * \log\left(\exp^{\log\left(p_1\right)+d_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+d_n}\right)+
 * \log\left(\exp^{\log\left(p_1\right)+f_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+f_n}\right)\right]
 * =\frac{e^{d_x}}{e^{d_1}p_1+\cdot\cdot\cdot+e^{d_m}p_m}+
 * \frac{e^{f_x}}{e^{f_1}p_1+\cdot\cdot\cdot+e^{f_m}p_m}
 * \f]
 *
 * \f[
 * \frac{\partial }{\partial d_x}\left[
 * \log\left(\exp^{\log\left(p_1\right)+d_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+d_n}\right)
 * +\log\left(\exp^{\log\left(p_1\right)+f_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+f_n}\right)\right]
 * =\frac{e^{d_x}p_x}{e^{d_1}p_1+\cdot\cdot\cdot+e^{d_m}p_m}
 * \f]
 *
 * @param theta vector of mixing proportions in [0, 1].
 * @param lambda array containing vectors of log densities.
 * @return log mixture of densities in specified proportion
 */
template <typename T_theta, typename T_lam, int R, int C>
typename return_type<T_theta, std::vector<Eigen::Matrix<T_lam, R, C> > >::type
log_mix(const T_theta& theta,
        const std::vector<Eigen::Matrix<T_lam, R, C> >& lambda) {
  static const char* function = "log_mix";
  typedef typename stan::partials_return_type<
```

```
      T_theta, std::vector<Eigen::Matrix<T_lam, R, C> > >::type
      T_partials_return;

  typedef typename Eigen::Matrix<T_partials_return, -1, 1> T_partials_vec;

  typedef typename Eigen::Matrix<T_partials_return, -1, -1> T_partials_mat;

  typedef typename std::vector<Eigen::Matrix<T_lam, R, C> > T_lamvec_type;

  const int N = length(lambda);
  const int M = theta.size();

  check_bounded(function, "theta", theta, 0, 1);
  check_not_nan(function, "theta", theta);
  check_finite(function, "theta", theta);
  for (int n = 0; n < N; ++n) {
    check_not_nan(function, "lambda", lambda[n]);
    check_finite(function, "lambda", lambda[n]);
    check_consistent_sizes(function, "theta", theta, "lambda", lambda[n]);
  }

  scalar_seq_view<T_theta> theta_vec(theta);
  T_partials_vec theta_dbl(M);
  for (int m = 0; m < M; ++m)
    theta_dbl[m] = value_of(theta_vec[m]);

  T_partials_mat lam_dbl(M, N);
  vector_seq_view<T_lamvec_type> lam_vec(lambda);
  for (int n = 0; n < N; ++n)
    for (int m = 0; m < M; ++m)
      lam_dbl(m, n) = value_of(lam_vec[n][m]);

  T_partials_mat logp_tmp = log(theta_dbl).replicate(1, N) + lam_dbl;

  T_partials_vec logp(N);
  for (int n = 0; n < N; ++n)
    logp[n] = log_sum_exp(logp_tmp.col(n).eval());

  operands_and_partials<T_theta, T_lamvec_type> ops_partials(theta, lambda);

  if (!(is_constant_struct<T_theta>::value
        && is_constant_struct<T_lam>::value)) {
    T_partials_mat derivs
        = (lam_dbl - logp.transpose().replicate(M, 1))
              .unaryExpr([](T_partials_return x) { return exp(x); });
    if (!is_constant_struct<T_theta>::value) {
      for (int m = 0; m < M; ++m)
        ops_partials.edge1_.partials_[m] = derivs.row(m).sum();
    }

    if (!is_constant_struct<T_lam>::value) {
      for (int n = 0; n < N; ++n)
        ops_partials.edge2_.partials_vec_[n]
            = derivs.col(n).cwiseProduct(theta_dbl);
    }
  }
  return ops_partials.build(logp.sum());
}
/**
 * Return the log mixture density given specified mixing proportions
 * and array of log density arrays.
 *
 * \f[
 * \frac{\partial }{\partial p_x}\left[
 * \log\left(\exp^{\log\left(p_1\right)+d_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+d_n}\right)+
 * \log\left(\exp^{\log\left(p_1\right)+f_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+f_n}\right)\right]
```

```
 * =\frac{e^{d_x}}{e^{d_1}p_1+\cdot\cdot\cdot+e^{d_m}p_m}+
 * \frac{e^{f_x}}{e^{f_1}p_1+\cdot\cdot\cdot+e^{f_m}p_m}
 * \f]
 *
 * \f[
 * \frac{\partial }{\partial d_x}\left[
 * \log\left(\exp^{\log\left(p_1\right)+d_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+d_n}\right)
 * +\log\left(\exp^{\log\left(p_1\right)+f_1}+\cdot\cdot\cdot+
 * \exp^{\log\left(p_n\right)+f_n}\right)\right]
 * =\frac{e^{d_x}p_x}{e^{d_1}p_1+\cdot\cdot\cdot+e^{d_m}p_m}
 * \f]
 *
 * @param theta vector of mixing proportions in [0, 1].
 * @param lambda array containing arrays of log densities.
 * @return log mixture of densities in specified proportion
 */
template <typename T_theta, typename T_lam>
typename return_type<T_theta, std::vector<std::vector<T_lam> > >::type log_mix(
    const T_theta& theta, const std::vector<std::vector<T_lam> >& lambda) {
  static const char* function = "log_mix";
  typedef typename stan::partials_return_type<
      T_theta, std::vector<std::vector<T_lam> > >::type T_partials_return;

  typedef typename Eigen::Matrix<T_partials_return, -1, 1> T_partials_vec;

  typedef typename Eigen::Matrix<T_partials_return, -1, -1> T_partials_mat;

  typedef typename std::vector<std::vector<T_lam> > T_lamvec_type;

  const int N = length(lambda);
  const int M = theta.size();

  check_bounded(function, "theta", theta, 0, 1);
  check_not_nan(function, "theta", theta);
  check_finite(function, "theta", theta);
  for (int n = 0; n < N; ++n) {
    check_not_nan(function, "lambda", lambda[n]);
    check_finite(function, "lambda", lambda[n]);
    check_consistent_sizes(function, "theta", theta, "lambda", lambda[n]);
  }

  scalar_seq_view<T_theta> theta_vec(theta);
  T_partials_vec theta_dbl(M);
  for (int m = 0; m < M; ++m)
    theta_dbl[m] = value_of(theta_vec[m]);

  T_partials_mat lam_dbl(M, N);
  for (int n = 0; n < N; ++n)
    for (int m = 0; m < M; ++m)
      lam_dbl(m, n) = value_of(lambda[n][m]);

  T_partials_mat logp_tmp = log(theta_dbl).replicate(1, N) + lam_dbl;

  T_partials_vec logp(N);
  for (int n = 0; n < N; ++n)
    logp[n] = log_sum_exp(logp_tmp.col(n).eval());

  T_partials_mat derivs
      = (lam_dbl - logp.transpose().replicate(M, 1))
            .unaryExpr([](T_partials_return x) { return exp(x); });

  T_partials_mat lam_deriv(M, N);
  for (int n = 0; n < N; ++n)
    lam_deriv.col(n) = derivs.col(n).cwiseProduct(theta_dbl);

  operands_and_partials<T_theta, T_lamvec_type> ops_partials(theta, lambda);
  if (!is_constant_struct<T_theta>::value) {
```

```
    for (int m = 0; m < M; ++m)
      ops_partials.edge1_.partials_[m] = derivs.row(m).sum();
  }

  if (!is_constant_struct<T_lam>::value) {
    for (int n = 0; n < N; ++n)
      for (int m = 0; m < M; ++m)
        ops_partials.edge2_.partials_vec_[n][m] = lam_deriv(m, n);
  }
  return ops_partials.build(logp.sum());
}
}  // namespace math
}  // namespace stan
#endif
```

10.5 Technical Appendix E – Ordered Logistic Log-Probability Function

```
#ifndef STAN_MATH_PRIM_MAT_PROB_ORDERED_LOGISTIC_LPMF_HPP
#define STAN_MATH_PRIM_MAT_PROB_ORDERED_LOGISTIC_LPMF_HPP

#include <stan/math/prim/mat/fun/value_of.hpp>
#include <stan/math/prim/mat/fun/size.hpp>
#include <stan/math/prim/mat/meta/vector_seq_view.hpp>
#include <stan/math/prim/mat/meta/length_mvt.hpp>
#include <stan/math/prim/mat/err/check_ordered.hpp>
#include <stan/math/prim/scal/fun/inv_logit.hpp>
#include <stan/math/prim/scal/fun/log1p_exp.hpp>
#include <stan/math/prim/scal/fun/log_inv_logit_diff.hpp>
#include <stan/math/prim/scal/fun/is_integer.hpp>
#include <stan/math/prim/scal/err/domain_error_vec.hpp>
#include <stan/math/prim/scal/err/check_bounded.hpp>
#include <stan/math/prim/scal/err/check_size_match.hpp>
#include <stan/math/prim/scal/err/check_finite.hpp>
#include <stan/math/prim/scal/err/check_greater.hpp>
#include <stan/math/prim/scal/err/check_consistent_sizes.hpp>
#include <stan/math/prim/scal/meta/include_summand.hpp>
#include <stan/math/prim/scal/meta/return_type.hpp>
#include <stan/math/prim/scal/meta/partials_return_type.hpp>
#include <stan/math/prim/scal/meta/operands_and_partials.hpp>
#include <stan/math/prim/scal/meta/is_constant_struct.hpp>
#include <stan/math/prim/scal/meta/scalar_seq_view.hpp>
#include <vector>

namespace stan {
namespace math {

/**
 * Returns the (natural) log probability of the specified array
 * of integers given the vector of continuous locations and
 * specified cutpoints in an ordered logistic model.
 *
 * <p>Typically the continous location
 * will be the dot product of a vector of regression coefficients
 * and a vector of predictors for the outcome
 *
  \f[
    \frac{\partial }{\partial \lambda} =
    \begin{cases}\\
    -\mathrm{logit}^{-1}(\lambda - c_1) & \mbox{if } k = 1,\\
    -(((1-e^{c_{k-1}-c_{k-2}})^{-1} - \mathrm{logit}^{-1}(c_{k-2}-\lambda)) +
    ((1-e^{c_{k-2}-c_{k-1}})^{-1} - \mathrm{logit}^{-1}(c_{k-1}-\lambda)))
    & \mathrm{if } 1 < k < K, \mathrm{and}\\
    \mathrm{logit}^{-1}(c_{K-2}-\lambda) & \mathrm{if } k = K.
    \end{cases}
  \f]

  \f[
    \frac{\partial }{\partial \lambda} =
    \begin{cases}
    -\mathrm{logit}^{-1}(\lambda - c_1) & \text{if } k = 1,\\
    -(((1-e^{c_{k-1}-c_{k-2}})^{-1} - \mathrm{logit}^{-1}(c_{k-2}-\lambda)) +
    ((1-e^{c_{k-2}-c_{k-1}})^{-1} - \mathrm{logit}^{-1}(c_{k-1}-\lambda)))
    & \text{if } 1 < k < K, \text{ and}\\
    \mathrm{logit}^{-1}(c_{K-2}-\lambda) & \text{if } k = K.
    \end{cases}
  \f]
 *
 * @tparam propto True if calculating up to a proportion.
 * @tparam T_y Y variable type (integer or array of integers).
 * @tparam T_loc Location type.
 * @tparam T_cut Cut-point type.
 * @param y Array of integers
```

```
 * @param lambda Vector of continuous location variables.
 * @param c Positive increasing vector of cutpoints.
 * @return Log probability of outcome given location and
 * cutpoints.
 * @throw std::domain_error If the outcome is not between 1 and
 * the number of cutpoints plus 2; if the cutpoint vector is
 * empty; if the cutpoint vector contains a non-positive,
 * non-finite value; or if the cutpoint vector is not sorted in
 * ascending order.
 * @throw std::invalid_argument If y and lambda are different
 * lengths.
 */
template <bool propto, typename T_y, typename T_loc, typename T_cut>
typename return_type<T_loc, T_cut>::type ordered_logistic_lpmf(
    const T_y& y, const T_loc& lambda, const T_cut& c) {
  static const char* function = "ordered_logistic";

  typedef
      typename stan::partials_return_type<T_loc, T_cut>::type T_partials_return;
  typedef typename Eigen::Matrix<T_partials_return, -1, 1> T_partials_vec;

  scalar_seq_view<T_loc> lam_vec(lambda);
  scalar_seq_view<T_y> y_vec(y);
  vector_seq_view<T_cut> c_vec(c);

  int K = c_vec[0].size() + 1;
  int N = length(lambda);
  int C_l = length_mvt(c);

  check_consistent_sizes(function, "Integers", y, "Locations", lambda);
  if (C_l > 1)
    check_size_match(function, "Length of location variables ", N,
                     "Number of cutpoint vectors ", C_l);

  int size_c_old = c_vec[0].size();
  for (int i = 1; i < C_l; i++) {
    int size_c_new = c_vec[i].size();

    check_size_match(function, "Size of one of the vectors of cutpoints ",
                     size_c_new, "Size of another vector of the cutpoints ",
                     size_c_old);
  }

  for (int n = 0; n < N; n++) {
    check_bounded(function, "Random variable", y_vec[n], 1, K);
    check_finite(function, "Location parameter", lam_vec[n]);
  }

  for (int i = 0; i < C_l; i++) {
    check_ordered(function, "Cut-points", c_vec[i]);
    check_greater(function, "Size of cut points parameter", c_vec[i].size(), 0);
    check_finite(function, "Final cut-point", c_vec[i](c_vec[i].size() - 1));
    check_finite(function, "First cut-point", c_vec[i](0));
  }

  operands_and_partials<T_loc, T_cut> ops_partials(lambda, c);

  T_partials_return logp(0.0);
  T_partials_vec c_dbl = value_of(c_vec[0]).template cast<T_partials_return>();

  for (int n = 0; n < N; ++n) {
    if (C_l > 1)
      c_dbl = value_of(c_vec[n]).template cast<T_partials_return>();
    T_partials_return lam_dbl = value_of(lam_vec[n]);

    if (y_vec[n] == 1) {
      logp -= log1p_exp(lam_dbl - c_dbl[0]);
      T_partials_return d = inv_logit(lam_dbl - c_dbl[0]);
```

```
      if (!is_constant_struct<T_loc>::value)
        ops_partials.edge1_.partials_[n] -= d;

      if (!is_constant_struct<T_cut>::value)
        ops_partials.edge2_.partials_vec_[n](0) = d;

    } else if (y_vec[n] == K) {
      logp -= log1p_exp(c_dbl[K - 2] - lam_dbl);
      T_partials_return d = inv_logit(c_dbl[K - 2] - lam_dbl);

      if (!is_constant_struct<T_loc>::value)
        ops_partials.edge1_.partials_[n] = d;

      if (!is_constant_struct<T_cut>::value)
        ops_partials.edge2_.partials_vec_[n](K - 2) -= d;

    } else {
      T_partials_return d1
          = inv(1 - exp(c_dbl[y_vec[n] - 1] - c_dbl[y_vec[n] - 2]))
            - inv_logit(c_dbl[y_vec[n] - 2] - lam_dbl);
      T_partials_return d2
          = inv(1 - exp(c_dbl[y_vec[n] - 2] - c_dbl[y_vec[n] - 1]))
            - inv_logit(c_dbl[y_vec[n] - 1] - lam_dbl);
      logp += log_inv_logit_diff(lam_dbl - c_dbl[y_vec[n] - 2],
                                 lam_dbl - c_dbl[y_vec[n] - 1]);

      if (!is_constant_struct<T_loc>::value)
        ops_partials.edge1_.partials_[n] -= d1 + d2;

      if (!is_constant_struct<T_cut>::value) {
        ops_partials.edge2_.partials_vec_[n](y_vec[n] - 2) += d1;
        ops_partials.edge2_.partials_vec_[n](y_vec[n] - 1) += d2;
      }
    }
  }
  return ops_partials.build(logp);
}

template <typename T_y, typename T_loc, typename T_cut>
typename return_type<T_loc, T_cut>::type ordered_logistic_lpmf(
    const T_y& y, const T_loc& lambda, const T_cut& c) {
  return ordered_logistic_lpmf<false>(y, lambda, c);
}

}  // namespace math
}  // namespace stan
#endif
```

11 Supplementary materials

11.1 Supplementary Material A – Study 2 Model 1 Stan Syntax

```
data{
  int N;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
}

transformed data{
//Factor Loadings
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int r_akin[3] = {8,10,12};
  int l_akin[3] = {9,11,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                  16,17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,
                   20,21,22,23,24,25,26,27,28,29,30,31,32,33};
  int r_akin_n[30] = {1,2,3,4,5,6,7,9,11,13,14,15,16,17,18,
                      19,20,21,22,23,24,25,26,27,28,29,30,31,32,33};
  int l_akin_n[30] = {1,2,3,4,5,6,7,8,10,12,14,15,16,17,18,19,20,21,
                      22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                   16,17,18,19,20,21,22,23,28,29,30,31,32,33};

  vector[25] ax_othl = rep_vector(0.0,25);
  vector[27] rt_othl = rep_vector(0.0,27);
  vector[28] rig_othl = rep_vector(0.0,28);
  vector[30] r_akin_othl = rep_vector(0.0,30);
  vector[30] l_akin_othl = rep_vector(0.0,30);
  vector[29] lakin_othl = rep_vector(0.0,29);
  vector[29] ktrem_othl = rep_vector(0.0,29);

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] eta_raw[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[3] r_akin_load;
```

```
  vector<lower=0.0>[3] l_akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[axial,1]    = ax_load;
  lam[rtrem,2]    = rt_load;
  lam[rigid,3]    = rig_load;
  lam[r_akin,4]   = r_akin_load;
  lam[l_akin,5]   = l_akin_load;
  lam[lakin,6]    = lakin_load;
  lam[ktrem,7]    = ktrem_load;

  lam[ax_n,1]     = ax_othl;
  lam[rt_n,2]     = rt_othl;
  lam[rig_n,3]    = rig_othl;
  lam[r_akin_n,4] = r_akin_othl;
  lam[l_akin_n,5] = l_akin_othl;
  lam[lak_n,6]    = lakin_othl;
  lam[ktr_n,7]    = ktrem_othl;

  lam_ecorr = lam * e_Lcorr;

  for(n in 1:N)
    y_star[n] = lam_ecorr * eta_raw[n] + y_star_raw[n];
}

model{
  e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  ax_load     ~ normal(load_mu,load_sigma);
  rt_load     ~ normal(load_mu,load_sigma);
  rig_load    ~ normal(load_mu,load_sigma);
  r_akin_load ~ normal(load_mu,load_sigma);
  l_akin_load ~ normal(load_mu,load_sigma);
  lakin_load  ~ normal(load_mu,load_sigma);
  ktrem_load  ~ normal(load_mu,load_sigma);

  for(n in 1:N) {
    y_star_raw[n] ~ std_normal();
    eta_raw[n]    ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}

generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] |  y_star[ii[m],jj[m]],
                                               thresholds[jj[m]]);
```

```
  deviance = -2.0 * sum(log_lik);
}
```

11.2 Supplementary Material B – Study 2 Model 2 Stan Syntax

```
data{
  int N;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
}

transformed data{
//Factor Loadings
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int r_akin[3] = {8,10,12};
  int l_akin[3] = {9,11,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,
                  15,16,17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int r_akin_n[30] = {1,2,3,4,5,6,7,9,11,13,14,15,16,17,18,19,20,21,
                      22,23,24,25,26,27,28,29,30,31,32,33};
  int l_akin_n[30] = {1,2,3,4,5,6,7,8,10,12,14,15,16,17,18,19,20,21,
                      22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                   17,18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] eta_raw[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[3] r_akin_load;
  vector<lower=0.0>[3] l_akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[30] r_akin_othl;
  vector[30] l_akin_othl;
  vector[29] lakin_othl;
```

```
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[axial,1]    = ax_load;
  lam[rtrem,2]    = rt_load;
  lam[rigid,3]    = rig_load;
  lam[r_akin,4]   = r_akin_load;
  lam[l_akin,5]   = l_akin_load;
  lam[lakin,6]    = lakin_load;
  lam[ktrem,7]    = ktrem_load;

  lam[ax_n,1]     = ax_othl;
  lam[rt_n,2]     = rt_othl;
  lam[rig_n,3]    = rig_othl;
  lam[r_akin_n,4] = r_akin_othl;
  lam[l_akin_n,5] = l_akin_othl;
  lam[lak_n,6]    = lakin_othl;
  lam[ktr_n,7]    = ktrem_othl;

  lam_ecorr = lam * e_Lcorr;

  for(n in 1:N) {
    y_star[n] = lam_ecorr * eta_raw[n] + y_star_raw[n];
  }
}

model{
  e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  ax_load     ~ normal(load_mu,load_sigma);
  rt_load     ~ normal(load_mu,load_sigma);
  rig_load    ~ normal(load_mu,load_sigma);
  r_akin_load ~ normal(load_mu,load_sigma);
  l_akin_load ~ normal(load_mu,load_sigma);
  lakin_load  ~ normal(load_mu,load_sigma);
  ktrem_load  ~ normal(load_mu,load_sigma);

  ax_othl     ~ normal(0,u_sd);
  rt_othl     ~ normal(0,u_sd);
  rig_othl    ~ normal(0,u_sd);
  r_akin_othl ~ normal(0,u_sd);
  l_akin_othl ~ normal(0,u_sd);
  lakin_othl  ~ normal(0,u_sd);
  ktrem_othl  ~ normal(0,u_sd);

  for(n in 1:N) {
    y_star_raw[n] ~ std_normal();
    eta_raw[n]    ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}
```

```
generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                           thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
}
```

11.3 Supplementary Material C – Study 2 Model 3 Stan Syntax

```
data{
  int N;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
}

transformed data{
//Factor Loadings
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int r_akin[3] = {8,10,12};
  int l_akin[3] = {9,11,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                  17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int r_akin_n[30] = {1,2,3,4,5,6,7,9,11,13,14,15,16,17,18,19,
                      20,21,22,23,24,25,26,27,28,29,30,31,32,33};
  int l_akin_n[30] = {1,2,3,4,5,6,7,8,10,12,14,15,16,17,18,19,20,21,
                      22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,22,
                   23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,
                   19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] eta_raw[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr;
  cholesky_factor_corr[J] u_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[3] r_akin_load;
  vector<lower=0.0>[3] l_akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[30] r_akin_othl;
  vector[30] l_akin_othl;
```

```
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[axial,1]  = ax_load;
  lam[rtrem,2]  = rt_load;
  lam[rigid,3]  = rig_load;
  lam[r_akin,4] = r_akin_load;
  lam[l_akin,5] = l_akin_load;
  lam[lakin,6]  = lakin_load;
  lam[ktrem,7]  = ktrem_load;

  lam[ax_n,1]     = ax_othl;
  lam[rt_n,2]     = rt_othl;
  lam[rig_n,3]    = rig_othl;
  lam[r_akin_n,4] = r_akin_othl;
  lam[l_akin_n,5] = l_akin_othl;
  lam[lak_n,6]    = lakin_othl;
  lam[ktr_n,7]    = ktrem_othl;

  lam_ecorr = lam * e_Lcorr;

  for(n in 1:N)
    y_star[n] = lam_ecorr * eta_raw[n] + u_Lcorr * y_star_raw[n];
}

model{
  e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  ax_load     ~ normal(load_mu,load_sigma);
  rt_load     ~ normal(load_mu,load_sigma);
  rig_load    ~ normal(load_mu,load_sigma);
  r_akin_load ~ normal(load_mu,load_sigma);
  l_akin_load ~ normal(load_mu,load_sigma);
  lakin_load  ~ normal(load_mu,load_sigma);
  ktrem_load  ~ normal(load_mu,load_sigma);

  ax_othl     ~ normal(0,u_sd);
  rt_othl     ~ normal(0,u_sd);
  rig_othl    ~ normal(0,u_sd);
  r_akin_othl ~ normal(0,u_sd);
  l_akin_othl ~ normal(0,u_sd);
  lakin_othl  ~ normal(0,u_sd);
  ktrem_othl  ~ normal(0,u_sd);

  u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N) {
    y_star_raw[n] ~ std_normal();
    eta_raw[n]    ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
```

```
}

generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                           thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
}
```

11.4 Supplementary Material D – Study 2 Model 4 Stan Syntax

```
data{
  int N;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
}

transformed data{
//Factor Loadings
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int r_akin[3] = {8,10,12};
  int l_akin[3] = {9,11,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
  int gen[33] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,
                 19,20,21,22,23,24,25,26,27,28,29,30,31,32,33};
//Cross Loadings
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                  17,24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                  17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int r_akin_n[30] = {1,2,3,4,5,6,7,9,11,13,14,15,16,17,18,19,20,
                      21,22,23,24,25,26,27,28,29,30,31,32,33};
  int l_akin_n[30] = {1,2,3,4,5,6,7,8,10,12,14,15,16,17,18,19,20,21,
                      22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                   17,18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[(E-1)] eta_s_raw[N];
  vector[N] eta_b_raw;
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E-1] e_Lcorr;
  cholesky_factor_corr[J] u_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[3] r_akin_load;
  vector<lower=0.0>[3] l_akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;
  vector[33] gen_load;

  vector[25] ax_othl;
```

```
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[30] r_akin_othl;
  vector[30] l_akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[axial,1]  = ax_load;
  lam[rtrem,2]  = rt_load;
  lam[rigid,3]  = rig_load;
  lam[r_akin,4] = r_akin_load;
  lam[l_akin,5] = l_akin_load;
  lam[lakin,6]  = lakin_load;
  lam[ktrem,7]  = ktrem_load;
  lam[gen,8]    = gen_load;

  lam[ax_n,1]     = ax_othl;
  lam[rt_n,2]     = rt_othl;
  lam[rig_n,3]    = rig_othl;
  lam[r_akin_n,4] = r_akin_othl;
  lam[l_akin_n,5] = l_akin_othl;
  lam[lak_n,6]    = lakin_othl;
  lam[ktr_n,7]    = ktrem_othl;

  lam_ecorr = lam*append_col(append_row(e_Lcorr,[0,0,0,0,0,0,0]),
                             [0,0,0,0,0,0,0,1]');

  for(n in 1:N)
    y_star[n] = (lam_ecorr*append_row(eta_s_raw[n],eta_b_raw[n]))
                + (u_Lcorr*y_star_raw[n]);
}

model{
  e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  ax_load     ~ normal(load_mu,load_sigma);
  rt_load     ~ normal(load_mu,load_sigma);
  rig_load    ~ normal(load_mu,load_sigma);
  r_akin_load ~ normal(load_mu,load_sigma);
  l_akin_load ~ normal(load_mu,load_sigma);
  lakin_load  ~ normal(load_mu,load_sigma);
  ktrem_load  ~ normal(load_mu,load_sigma);
  gen_load    ~ normal(load_mu,load_sigma);

  ax_othl     ~ normal(0,u_sd);
  rt_othl     ~ normal(0,u_sd);
  rig_othl    ~ normal(0,u_sd);
  r_akin_othl ~ normal(0,u_sd);
  l_akin_othl ~ normal(0,u_sd);
  lakin_othl  ~ normal(0,u_sd);
  ktrem_othl  ~ normal(0,u_sd);

  u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N) {
    eta_s_raw[n]  ~ std_normal();
    y_star_raw[n] ~ std_normal();
  }
```

```
  eta_b_raw ~ std_normal();

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}

generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                             thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
}
```

11.5 Supplementary Material E – Study 2 Model 5 Stan Syntax

```
data{
  int N;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y[M];
}

transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int r_akin[3] = {8,10,12};
  int l_akin[3] = {9,11,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int r_akin_n[30] = {1,2,3,4,5,6,7,9,11,13,14,15,16,17,18,19,
                      20,21,22,23,24,25,26,27,28,29,30,31,32,33};
  int l_akin_n[30] = {1,2,3,4,5,6,7,8,10,12,14,15,16,17,18,19,20,
                      21,22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                   18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] eta_raw[N];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] e_Lcorr;
  cholesky_factor_corr[J] u_Lcorr;
  vector[J] y_star_raw[N];

  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[3] r_akin_load;
  vector<lower=0.0>[3] l_akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;
```

```
  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[30] r_akin_othl;
  vector[30] l_akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] lam_ecorr;
  vector[J] y_star[N];

  lam[right,1]  = right_load;
  lam[left,2]   = left_load;
  lam[axial,3]  = ax_load;
  lam[rtrem,4]  = rt_load;
  lam[rigid,5]  = rig_load;
  lam[r_akin,6] = r_akin_load;
  lam[l_akin,7] = l_akin_load;
  lam[lakin,8]  = lakin_load;
  lam[ktrem,9]  = ktrem_load;

  lam[r_n,1]      = right_othl;
  lam[l_n,2]      = left_othl;
  lam[ax_n,3]     = ax_othl;
  lam[rt_n,4]     = rt_othl;
  lam[rig_n,5]    = rig_othl;
  lam[r_akin_n,6] = r_akin_othl;
  lam[l_akin_n,7] = l_akin_othl;
  lam[lak_n,8]    = lakin_othl;
  lam[ktr_n,9]    = ktrem_othl;

  lam_ecorr = lam * e_Lcorr;

  for(n in 1:N)
    y_star[n] = lam_ecorr * eta_raw[n] + u_Lcorr * y_star_raw[n];
}

model{
  e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load  ~ normal(load_mu,load_sigma);
  left_load   ~ normal(load_mu,load_sigma);
  ax_load     ~ normal(load_mu,load_sigma);
  rt_load     ~ normal(load_mu,load_sigma);
  rig_load    ~ normal(load_mu,load_sigma);
  r_akin_load ~ normal(load_mu,load_sigma);
  l_akin_load ~ normal(load_mu,load_sigma);
  lakin_load  ~ normal(load_mu,load_sigma);
  ktrem_load  ~ normal(load_mu,load_sigma);

  right_othl  ~ normal(0,u_sd);
  left_othl   ~ normal(0,u_sd);
  ax_othl     ~ normal(0,u_sd);
  rt_othl     ~ normal(0,u_sd);
  rig_othl    ~ normal(0,u_sd);
  r_akin_othl ~ normal(0,u_sd);
  l_akin_othl ~ normal(0,u_sd);
  lakin_othl  ~ normal(0,u_sd);
  ktrem_othl  ~ normal(0,u_sd);
```

```
  u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N) {
    y_star_raw[n] ~ std_normal();
    eta_raw[n]    ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y[m] ~ ordered_logistic(y_star[ii[m],jj[m]],thresholds[jj[m]]);
}

generated quantities {
  vector[M] log_lik;
  real deviance;

  for(m in 1:M)
    log_lik[m] = ordered_logistic_lpmf(y[m] | y_star[ii[m],jj[m]],
                                           thresholds[jj[m]]);

  deviance = -2.0 * sum(log_lik);
}
```

11.6 Supplementary Material F – Study 2 Invariance Model: All Free Stan Syntax

```
data{
  int N_pc;
  int N_pp;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y_pc[M];
  int y_pp[N_pp,J];
}

transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int akin[6] = {8,9,10,11,12,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                  16,17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,20,21,
                    22,23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                   17,18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] pc_eta_raw[N_pc];
  vector[E] pp_eta_raw[N_pp];
  real pc_thresh_mu;
  real pp_thresh_mu;
  real<lower=0> pc_thresh_sigma;
  real<lower=0> pp_thresh_sigma;
  real pc_load_mu;
  real pp_load_mu;
  real<lower=0> pc_load_sigma;
  real<lower=0> pp_load_sigma;
  ordered[T] pc_thresholds[J];
  ordered[T] pp_thresholds[J];
  cholesky_factor_corr[E] pc_e_Lcorr;
  cholesky_factor_corr[E] pp_e_Lcorr;
  cholesky_factor_corr[J] pc_u_Lcorr;
  cholesky_factor_corr[J] pp_u_Lcorr;
  vector[J] pc_ystar_raw[N_pc];
  vector[J] pp_ystar_raw[N_pp];
```

```
  vector<lower=0.0>[11] pc_right_load;
  vector<lower=0.0>[11] pc_left_load;
  vector<lower=0.0>[8] pc_ax_load;
  vector<lower=0.0>[6] pc_rt_load;
  vector<lower=0.0>[5] pc_rig_load;
  vector<lower=0.0>[6] pc_akin_load;
  vector<lower=0.0>[4] pc_lakin_load;
  vector<lower=0.0>[4] pc_ktrem_load;

  vector[22] pc_right_othl;
  vector[22] pc_left_othl;
  vector[25] pc_ax_othl;
  vector[27] pc_rt_othl;
  vector[28] pc_rig_othl;
  vector[27] pc_akin_othl;
  vector[29] pc_lakin_othl;
  vector[29] pc_ktrem_othl;


  vector<lower=0.0>[11] pp_right_load;
  vector<lower=0.0>[11] pp_left_load;
  vector<lower=0.0>[8] pp_ax_load;
  vector<lower=0.0>[6] pp_rt_load;
  vector<lower=0.0>[5] pp_rig_load;
  vector<lower=0.0>[6] pp_akin_load;
  vector<lower=0.0>[4] pp_lakin_load;
  vector<lower=0.0>[4] pp_ktrem_load;

  vector[22] pp_right_othl;
  vector[22] pp_left_othl;
  vector[25] pp_ax_othl;
  vector[27] pp_rt_othl;
  vector[28] pp_rig_othl;
  vector[27] pp_akin_othl;
  vector[29] pp_lakin_othl;
  vector[29] pp_ktrem_othl;
}

transformed parameters{
  matrix[J,E] pc_lam;
  matrix[J,E] pp_lam;
  matrix[J,E] pc_lam_ecorr;
  matrix[J,E] pp_lam_ecorr;
  vector[J] pc_ystar[N_pc];
  vector[J] pp_ystar[N_pp];

  pc_lam[right,1] = pc_right_load;
  pc_lam[left,2]  = pc_left_load;
  pc_lam[axial,3] = pc_ax_load;
  pc_lam[rtrem,4] = pc_rt_load;
  pc_lam[rigid,5] = pc_rig_load;
  pc_lam[akin,6]  = pc_akin_load;
  pc_lam[lakin,7] = pc_lakin_load;
  pc_lam[ktrem,8] = pc_ktrem_load;

  pc_lam[r_n,1]    = pc_right_othl;
  pc_lam[l_n,2]    = pc_left_othl;
  pc_lam[ax_n,3]   = pc_ax_othl;
  pc_lam[rt_n,4]   = pc_rt_othl;
  pc_lam[rig_n,5]  = pc_rig_othl;
  pc_lam[akin_n,6] = pc_akin_othl;
  pc_lam[lak_n,7]  = pc_lakin_othl;
  pc_lam[ktr_n,8]  = pc_ktrem_othl;

  pp_lam[right,1] = pp_right_load;
  pp_lam[left,2]  = pp_left_load;
  pp_lam[axial,3] = pp_ax_load;
```

```
  pp_lam[rtrem,4] = pp_rt_load;
  pp_lam[rigid,5] = pp_rig_load;
  pp_lam[akin,6]  = pp_akin_load;
  pp_lam[lakin,7] = pp_lakin_load;
  pp_lam[ktrem,8] = pp_ktrem_load;

  pp_lam[r_n,1]    = pp_right_othl;
  pp_lam[l_n,2]    = pp_left_othl;
  pp_lam[ax_n,3]   = pp_ax_othl;
  pp_lam[rt_n,4]   = pp_rt_othl;
  pp_lam[rig_n,5]  = pp_rig_othl;
  pp_lam[akin_n,6] = pp_akin_othl;
  pp_lam[lak_n,7]  = pp_lakin_othl;
  pp_lam[ktr_n,8]  = pp_ktrem_othl;

  pc_lam_ecorr = pc_lam * pc_e_Lcorr;
  pp_lam_ecorr = pp_lam * pp_e_Lcorr;

  for(n in 1:N_pc)
    pc_ystar[n] = pc_lam_ecorr * pc_eta_raw[n] + pc_u_Lcorr * pc_ystar_raw[n];

  for(n in 1:N_pp)
    pp_ystar[n] = pp_lam_ecorr * pp_eta_raw[n] + pp_u_Lcorr * pp_ystar_raw[n];
}

model{
  pc_e_Lcorr ~ lkj_corr_cholesky(1);
  pp_e_Lcorr ~ lkj_corr_cholesky(1);

  pc_load_mu    ~ normal(0,5);
  pp_load_mu    ~ normal(0,5);
  pc_load_sigma ~ cauchy(0,5);
  pp_load_sigma ~ cauchy(0,5);

  pc_right_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_left_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_ax_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_rt_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_rig_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_akin_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_lakin_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_ktrem_load ~ normal(pc_load_mu,pc_load_sigma);

  pc_right_othl ~ normal(0,u_sd);
  pc_left_othl  ~ normal(0,u_sd);
  pc_ax_othl    ~ normal(0,u_sd);
  pc_rt_othl    ~ normal(0,u_sd);
  pc_rig_othl   ~ normal(0,u_sd);
  pc_akin_othl  ~ normal(0,u_sd);
  pc_lakin_othl ~ normal(0,u_sd);
  pc_ktrem_othl ~ normal(0,u_sd);

  pp_right_load ~ normal(pp_load_mu,pp_load_sigma);
  pp_left_load  ~ normal(pp_load_mu,pp_load_sigma);
  pp_ax_load    ~ normal(pp_load_mu,pp_load_sigma);
  pp_rt_load    ~ normal(pp_load_mu,pp_load_sigma);
  pp_rig_load   ~ normal(pp_load_mu,pp_load_sigma);
  pp_akin_load  ~ normal(pp_load_mu,pp_load_sigma);
  pp_lakin_load ~ normal(pp_load_mu,pp_load_sigma);
  pp_ktrem_load ~ normal(pp_load_mu,pp_load_sigma);

  pp_right_othl ~ normal(0,u_sd);
  pp_left_othl  ~ normal(0,u_sd);
  pp_ax_othl    ~ normal(0,u_sd);
  pp_rt_othl    ~ normal(0,u_sd);
  pp_rig_othl   ~ normal(0,u_sd);
  pp_akin_othl  ~ normal(0,u_sd);
  pp_lakin_othl ~ normal(0,u_sd);
```

```
  pp_ktrem_othl ~ normal(0,u_sd);

  pc_u_Lcorr ~ lkj_corr_cholesky(20);
  pp_u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N_pc) {
    pc_eta_raw[n]   ~ std_normal();
    pc_ystar_raw[n] ~ std_normal();
  }

  for(n in 1:N_pp) {
    pp_eta_raw[n]   ~ std_normal();
    pp_ystar_raw[n] ~ std_normal();
  }

  pc_thresh_mu    ~ normal(0,5);
  pp_thresh_mu    ~ normal(0,5);
  pc_thresh_sigma ~ cauchy(0,5);
  pp_thresh_sigma ~ cauchy(0,5);

  for(j in 1:J) {
    pc_thresholds[j] ~ normal(pc_thresh_mu,pc_thresh_sigma);
    pp_thresholds[j] ~ normal(pp_thresh_mu,pp_thresh_sigma);
  }

  for(m in 1:M)
    y_pc[m] ~ ordered_logistic(pc_ystar[ii[m],jj[m]],pc_thresholds[jj[m]]);

  for(n in 1:N_pp)
    y_pp[n] ~ ordered_logistic(pp_ystar[n],pp_thresholds);
}

generated quantities {
  matrix[N_pp,J] pp_ll_tmp;
  vector[M] pc_ll_tmp;
  vector[M + (N_pp*J)] log_lik;
  real deviance;

  for(m in 1:M)
    pc_ll_tmp[m] = ordered_logistic_lpmf(y_pc[m]| pc_ystar[ii[m],jj[m]],
                                                  pc_thresholds[jj[m]]);

  for(n in 1:N_pp)
    for(j in 1:J)
      pp_ll_tmp[n,j] = ordered_logistic_lpmf(y_pp[n,j]| pp_ystar[n,j],
                                                        pp_thresholds[j]);

  log_lik = append_row(pc_ll_tmp,to_vector(pp_ll_tmp));
  deviance = -2.0 * sum(log_lik);
}
```

11.7 Supplementary Material G – Study 2 Invariance Model: Loadings Invariant Stan Syntax

```
data{
  int N_pc;
  int N_pp;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y_pc[M];
  int y_pp[N_pp,J];
}

transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int akin[6] = {8,9,10,11,12,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                  17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,20,21,22,
                    23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,21,
                   22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                   18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] pc_eta_raw[N_pc];
  vector[E] pp_eta_raw[N_pp];
  real pc_thresh_mu;
  real pp_thresh_mu;
  real<lower=0> pc_thresh_sigma;
  real<lower=0> pp_thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] pc_thresholds[J];
  ordered[T] pp_thresholds[J];
  cholesky_factor_corr[E] pc_e_Lcorr;
  cholesky_factor_corr[E] pp_e_Lcorr;
  cholesky_factor_corr[J] pc_u_Lcorr;
  cholesky_factor_corr[J] pp_u_Lcorr;
  vector[J] pc_ystar_raw[N_pc];
  vector[J] pp_ystar_raw[N_pp];
```

```
  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[6] akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[27] akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] pc_lam_ecorr;
  matrix[J,E] pp_lam_ecorr;
  vector[J] pc_ystar[N_pc];
  vector[J] pp_ystar[N_pp];

  lam[right,1] = right_load;
  lam[left,2]  = left_load;
  lam[axial,3] = ax_load;
  lam[rtrem,4] = rt_load;
  lam[rigid,5] = rig_load;
  lam[akin,6]  = akin_load;
  lam[lakin,7] = lakin_load;
  lam[ktrem,8] = ktrem_load;

  lam[r_n,1]    = right_othl;
  lam[l_n,2]    = left_othl;
  lam[ax_n,3]   = ax_othl;
  lam[rt_n,4]   = rt_othl;
  lam[rig_n,5]  = rig_othl;
  lam[akin_n,6] = akin_othl;
  lam[lak_n,7]  = lakin_othl;
  lam[ktr_n,8]  = ktrem_othl;

  pc_lam_ecorr = lam * pc_e_Lcorr;
  pp_lam_ecorr = lam * pp_e_Lcorr;

  for(n in 1:N_pc)
    pc_ystar[n] = pc_lam_ecorr * pc_eta_raw[n] + pc_u_Lcorr * pc_ystar_raw[n];

  for(n in 1:N_pp)
    pp_ystar[n] = pp_lam_ecorr * pp_eta_raw[n] + pp_u_Lcorr * pp_ystar_raw[n];
}

model{
  pc_e_Lcorr ~ lkj_corr_cholesky(1);
  pp_e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load ~ normal(load_mu,load_sigma);
  left_load  ~ normal(load_mu,load_sigma);
  ax_load    ~ normal(load_mu,load_sigma);
  rt_load    ~ normal(load_mu,load_sigma);
  rig_load   ~ normal(load_mu,load_sigma);
  akin_load  ~ normal(load_mu,load_sigma);
```

```
  lakin_load ~ normal(load_mu,load_sigma);
  ktrem_load ~ normal(load_mu,load_sigma);

  right_othl ~ normal(0,u_sd);
  left_othl  ~ normal(0,u_sd);
  ax_othl    ~ normal(0,u_sd);
  rt_othl    ~ normal(0,u_sd);
  rig_othl   ~ normal(0,u_sd);
  akin_othl  ~ normal(0,u_sd);
  lakin_othl ~ normal(0,u_sd);
  ktrem_othl ~ normal(0,u_sd);

  pc_u_Lcorr ~ lkj_corr_cholesky(20);
  pp_u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N_pc) {
    pc_eta_raw[n]   ~ std_normal();
    pc_ystar_raw[n] ~ std_normal();
  }

  for(n in 1:N_pp) {
    pp_eta_raw[n]   ~ std_normal();
    pp_ystar_raw[n] ~ std_normal();
  }

  pc_thresh_mu    ~ normal(0,5);
  pp_thresh_mu    ~ normal(0,5);
  pc_thresh_sigma ~ cauchy(0,5);
  pp_thresh_sigma ~ cauchy(0,5);

  for(j in 1:J) {
    pc_thresholds[j] ~ normal(pc_thresh_mu,pc_thresh_sigma);
    pp_thresholds[j] ~ normal(pp_thresh_mu,pp_thresh_sigma);
  }

  for(m in 1:M)
    y_pc[m] ~ ordered_logistic(pc_ystar[ii[m],jj[m]],pc_thresholds[jj[m]]);

  for(n in 1:N_pp)
    y_pp[n] ~ ordered_logistic(pp_ystar[n],pp_thresholds);
}

generated quantities {
  matrix[N_pp,J] pp_ll_tmp;
  vector[M] pc_ll_tmp;
  vector[M + (N_pp*J)] log_lik;
  real deviance;

  for(m in 1:M)
    pc_ll_tmp[m] = ordered_logistic_lpmf(y_pc[m]| pc_ystar[ii[m],jj[m]],
                                                  pc_thresholds[jj[m]]);

  for(n in 1:N_pp)
    for(j in 1:J)
      pp_ll_tmp[n,j] = ordered_logistic_lpmf(y_pp[n,j] | pp_ystar[n,j],
                                                         pp_thresholds[j]);

  log_lik = append_row(pc_ll_tmp,to_vector(pp_ll_tmp));
  deviance = -2.0 * sum(log_lik);
}
```

11.8 Supplementary Material H – Study 2 Invariance: Thresholds Invariant Model Stan Syntax

```
data{
  int N_pc;
  int N_pp;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y_pc[M];
  int y_pp[N_pp,J];
}

transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int akin[6] = {8,9,10,11,12,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
                  24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                  16,17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,20,21,22,
                    23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                   17,18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
}

parameters{
  vector[E] pc_eta_raw[N_pc];
  vector[E] pp_eta_raw[N_pp];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real pc_load_mu;
  real pp_load_mu;
  real<lower=0> pc_load_sigma;
  real<lower=0> pp_load_sigma;
  ordered[T] thresholds[J];
  cholesky_factor_corr[E] pc_e_Lcorr;
  cholesky_factor_corr[E] pp_e_Lcorr;
  cholesky_factor_corr[J] pc_u_Lcorr;
  cholesky_factor_corr[J] pp_u_Lcorr;
  vector[J] pc_ystar_raw[N_pc];
  vector[J] pp_ystar_raw[N_pp];
```

```
  vector<lower=0.0>[11] pc_right_load;
  vector<lower=0.0>[11] pc_left_load;
  vector<lower=0.0>[8] pc_ax_load;
  vector<lower=0.0>[6] pc_rt_load;
  vector<lower=0.0>[5] pc_rig_load;
  vector<lower=0.0>[6] pc_akin_load;
  vector<lower=0.0>[4] pc_lakin_load;
  vector<lower=0.0>[4] pc_ktrem_load;

  vector[22] pc_right_othl;
  vector[22] pc_left_othl;
  vector[25] pc_ax_othl;
  vector[27] pc_rt_othl;
  vector[28] pc_rig_othl;
  vector[27] pc_akin_othl;
  vector[29] pc_lakin_othl;
  vector[29] pc_ktrem_othl;

  vector<lower=0.0>[11] pp_right_load;
  vector<lower=0.0>[11] pp_left_load;
  vector<lower=0.0>[8] pp_ax_load;
  vector<lower=0.0>[6] pp_rt_load;
  vector<lower=0.0>[5] pp_rig_load;
  vector<lower=0.0>[6] pp_akin_load;
  vector<lower=0.0>[4] pp_lakin_load;
  vector<lower=0.0>[4] pp_ktrem_load;

  vector[22] pp_right_othl;
  vector[22] pp_left_othl;
  vector[25] pp_ax_othl;
  vector[27] pp_rt_othl;
  vector[28] pp_rig_othl;
  vector[27] pp_akin_othl;
  vector[29] pp_lakin_othl;
  vector[29] pp_ktrem_othl;
}

transformed parameters{
  matrix[J,E] pc_lam;
  matrix[J,E] pp_lam;
  matrix[J,E] pc_lam_ecorr;
  matrix[J,E] pp_lam_ecorr;
  vector[J] pc_ystar[N_pc];
  vector[J] pp_ystar[N_pp];

  pc_lam[right,1] = pc_right_load;
  pc_lam[left,2]  = pc_left_load;
  pc_lam[axial,3] = pc_ax_load;
  pc_lam[rtrem,4] = pc_rt_load;
  pc_lam[rigid,5] = pc_rig_load;
  pc_lam[akin,6]  = pc_akin_load;
  pc_lam[lakin,7] = pc_lakin_load;
  pc_lam[ktrem,8] = pc_ktrem_load;

  pc_lam[r_n,1]    = pc_right_othl;
  pc_lam[l_n,2]    = pc_left_othl;
  pc_lam[ax_n,3]   = pc_ax_othl;
  pc_lam[rt_n,4]   = pc_rt_othl;
  pc_lam[rig_n,5]  = pc_rig_othl;
  pc_lam[akin_n,6] = pc_akin_othl;
  pc_lam[lak_n,7]  = pc_lakin_othl;
  pc_lam[ktr_n,8]  = pc_ktrem_othl;

  pp_lam[right,1] = pp_right_load;
  pp_lam[left,2]  = pp_left_load;
  pp_lam[axial,3] = pp_ax_load;
  pp_lam[rtrem,4] = pp_rt_load;
  pp_lam[rigid,5] = pp_rig_load;
```

```
  pp_lam[akin,6]  = pp_akin_load;
  pp_lam[lakin,7] = pp_lakin_load;
  pp_lam[ktrem,8] = pp_ktrem_load;
  pp_lam[r_n,1]    = pp_right_othl;
  pp_lam[l_n,2]    = pp_left_othl;
  pp_lam[ax_n,3]   = pp_ax_othl;
  pp_lam[rt_n,4]   = pp_rt_othl;
  pp_lam[rig_n,5]  = pp_rig_othl;
  pp_lam[akin_n,6] = pp_akin_othl;
  pp_lam[lak_n,7]  = pp_lakin_othl;
  pp_lam[ktr_n,8]  = pp_ktrem_othl;

  pc_lam_ecorr = pc_lam * pc_e_Lcorr;
  pp_lam_ecorr = pp_lam * pp_e_Lcorr;

  for(n in 1:N_pc)
    pc_ystar[n] = pc_lam_ecorr * pc_eta_raw[n] + pc_u_Lcorr * pc_ystar_raw[n];

  for(n in 1:N_pp)
    pp_ystar[n] = pp_lam_ecorr * pp_eta_raw[n] + pp_u_Lcorr * pp_ystar_raw[n];
}

model{
  pc_e_Lcorr ~ lkj_corr_cholesky(1);
  pp_e_Lcorr ~ lkj_corr_cholesky(1);

  pc_load_mu    ~ normal(0,5);
  pp_load_mu    ~ normal(0,5);
  pc_load_sigma ~ cauchy(0,5);
  pp_load_sigma ~ cauchy(0,5);

  pc_right_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_left_load  ~ normal(pc_load_mu,pc_load_sigma);
  pc_ax_load    ~ normal(pc_load_mu,pc_load_sigma);
  pc_rt_load    ~ normal(pc_load_mu,pc_load_sigma);
  pc_rig_load   ~ normal(pc_load_mu,pc_load_sigma);
  pc_akin_load  ~ normal(pc_load_mu,pc_load_sigma);
  pc_lakin_load ~ normal(pc_load_mu,pc_load_sigma);
  pc_ktrem_load ~ normal(pc_load_mu,pc_load_sigma);

  pc_right_othl ~ normal(0,u_sd);
  pc_left_othl  ~ normal(0,u_sd);
  pc_ax_othl    ~ normal(0,u_sd);
  pc_rt_othl    ~ normal(0,u_sd);
  pc_rig_othl   ~ normal(0,u_sd);
  pc_akin_othl  ~ normal(0,u_sd);
  pc_lakin_othl ~ normal(0,u_sd);
  pc_ktrem_othl ~ normal(0,u_sd);

  pp_right_load ~ normal(pp_load_mu,pp_load_sigma);
  pp_left_load  ~ normal(pp_load_mu,pp_load_sigma);
  pp_ax_load    ~ normal(pp_load_mu,pp_load_sigma);
  pp_rt_load    ~ normal(pp_load_mu,pp_load_sigma);
  pp_rig_load   ~ normal(pp_load_mu,pp_load_sigma);
  pp_akin_load  ~ normal(pp_load_mu,pp_load_sigma);
  pp_lakin_load ~ normal(pp_load_mu,pp_load_sigma);
  pp_ktrem_load ~ normal(pp_load_mu,pp_load_sigma);

  pp_right_othl ~ normal(0,u_sd);
  pp_left_othl  ~ normal(0,u_sd);
  pp_ax_othl    ~ normal(0,u_sd);
  pp_rt_othl    ~ normal(0,u_sd);
  pp_rig_othl   ~ normal(0,u_sd);
  pp_akin_othl  ~ normal(0,u_sd);
  pp_lakin_othl ~ normal(0,u_sd);
  pp_ktrem_othl ~ normal(0,u_sd);

  pc_u_Lcorr ~ lkj_corr_cholesky(20);
```

```
  pp_u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N_pc) {
    pc_eta_raw[n]   ~ std_normal();
    pc_ystar_raw[n] ~ std_normal();
  }

  for(n in 1:N_pp) {
    pp_eta_raw[n]   ~ std_normal();
    pp_ystar_raw[n] ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J)
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);

  for(m in 1:M)
    y_pc[m] ~ ordered_logistic(pc_ystar[ii[m],jj[m]],thresholds[jj[m]]);

  for(n in 1:N_pp)
    y_pp[n] ~ ordered_logistic(pp_ystar[n],thresholds);
}

generated quantities {
  matrix[N_pp,J] pp_ll_tmp;
  vector[M] pc_ll_tmp;
  vector[M + (N_pp*J)] log_lik;
  real deviance;

  for(m in 1:M)
    pc_ll_tmp[m] = ordered_logistic_lpmf(y_pc[m]| pc_ystar[ii[m],jj[m]],
                                                  thresholds[jj[m]]);

  for(n in 1:N_pp)
    for(j in 1:J)
      pp_ll_tmp[n,j] = ordered_logistic_lpmf(y_pp[n,j]| pp_ystar[n,j],
                                                        thresholds[j]);

  log_lik = append_row(pc_ll_tmp,to_vector(pp_ll_tmp));
  deviance = -2.0 * sum(log_lik);
}
```

11.9 Supplementary Material I – Study 2 Invariance: Loadings Invariant and Thresholds Approximately Invariant Stan Syntax

```
data{
  int N_pc;
  int N_pp;
  int J;
  int T;
  int E;
  int M;
  int ii[M];
  int jj[M];
  int y_pc[M];
  int y_pp[N_pp,J];
}

transformed data{
//Factor Loadings
  int right[11] = {4,6,8,10,12,14,16,24,26,28,30};
  int left[11] = {5,7,9,11,13,15,17,25,27,29,31};
  int axial[8] = {1,2,18,19,20,21,22,23};
  int rtrem[6] = {28,29,30,31,32,33};
  int rigid[5] = {3,4,5,6,7};
  int akin[6] = {8,9,10,11,12,13};
  int lakin[4] = {14,15,16,17};
  int ktrem[4] = {24,25,26,27};
//Cross Loadings
  int r_n[22] = {1,2,3,5,7,9,11,13,15,17,18,19,20,21,22,23,25,27,29,31,32,33};
  int l_n[22] = {1,2,3,4,6,8,10,12,14,16,18,19,20,21,22,23,24,26,28,30,32,33};
  int ax_n[25] = {3,4,5,6,7,8,9,10,11,12,13,14,15,16,
                  17,24,25,26,27,28,29,30,31,32,33};
  int rt_n[27] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                  16,17,18,19,20,21,22,23,24,25,26,27};
  int rig_n[28] = {1,2,8,9,10,11,12,13,14,15,16,17,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int akin_n[27] = {1,2,3,4,5,6,7,14,15,16,17,18,19,20,21,22,
                    23,24,25,26,27,28,29,30,31,32,33};
  int lak_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,18,19,20,
                   21,22,23,24,25,26,27,28,29,30,31,32,33};
  int ktr_n[29] = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,
                   16,17,18,19,20,21,22,23,28,29,30,31,32,33};

  real i_sd = 5.0;
  real u_sd = 0.1;
  real thr_diff = sqrt(0.5);
}

parameters{
  vector[E] pc_eta_raw[N_pc];
  vector[E] pp_eta_raw[N_pp];
  real thresh_mu;
  real<lower=0> thresh_sigma;
  real load_mu;
  real<lower=0> load_sigma;
  ordered[T] thresholds[J];
  ordered[T] pp_thresholds[J];
  cholesky_factor_corr[E] pc_e_Lcorr;
  cholesky_factor_corr[E] pp_e_Lcorr;
  cholesky_factor_corr[J] pc_u_Lcorr;
  cholesky_factor_corr[J] pp_u_Lcorr;
  vector[J] pc_ystar_raw[N_pc];
  vector[J] pp_ystar_raw[N_pp];
```

```
  vector<lower=0.0>[11] right_load;
  vector<lower=0.0>[11] left_load;
  vector<lower=0.0>[8] ax_load;
  vector<lower=0.0>[6] rt_load;
  vector<lower=0.0>[5] rig_load;
  vector<lower=0.0>[6] akin_load;
  vector<lower=0.0>[4] lakin_load;
  vector<lower=0.0>[4] ktrem_load;

  vector[22] right_othl;
  vector[22] left_othl;
  vector[25] ax_othl;
  vector[27] rt_othl;
  vector[28] rig_othl;
  vector[27] akin_othl;
  vector[29] lakin_othl;
  vector[29] ktrem_othl;
}

transformed parameters{
  matrix[J,E] lam;
  matrix[J,E] pc_lam_ecorr;
  matrix[J,E] pp_lam_ecorr;
  vector[J] pc_ystar[N_pc];
  vector[J] pp_ystar[N_pp];

  lam[right,1] = right_load;
  lam[left,2]  = left_load;
  lam[axial,3] = ax_load;
  lam[rtrem,4] = rt_load;
  lam[rigid,5] = rig_load;
  lam[akin,6]  = akin_load;
  lam[lakin,7] = lakin_load;
  lam[ktrem,8] = ktrem_load;

  lam[r_n,1]    = right_othl;
  lam[l_n,2]    = left_othl;
  lam[ax_n,3]   = ax_othl;
  lam[rt_n,4]   = rt_othl;
  lam[rig_n,5]  = rig_othl;
  lam[akin_n,6] = akin_othl;
  lam[lak_n,7]  = lakin_othl;
  lam[ktr_n,8]  = ktrem_othl;

  pc_lam_ecorr = lam * pc_e_Lcorr;
  pp_lam_ecorr = lam * pp_e_Lcorr;

  for(n in 1:N_pc)
    pc_ystar[n] = pc_lam_ecorr * pc_eta_raw[n] + pc_u_Lcorr * pc_ystar_raw[n];

  for(n in 1:N_pp)
    pp_ystar[n] = pp_lam_ecorr * pp_eta_raw[n] + pp_u_Lcorr * pp_ystar_raw[n];
}

model{
  pc_e_Lcorr ~ lkj_corr_cholesky(1);
  pp_e_Lcorr ~ lkj_corr_cholesky(1);

  load_mu    ~ normal(0,5);
  load_sigma ~ cauchy(0,5);

  right_load ~ normal(load_mu,load_sigma);
  left_load  ~ normal(load_mu,load_sigma);
  ax_load    ~ normal(load_mu,load_sigma);
  rt_load    ~ normal(load_mu,load_sigma);
  rig_load   ~ normal(load_mu,load_sigma);
  akin_load  ~ normal(load_mu,load_sigma);
  lakin_load ~ normal(load_mu,load_sigma);
```

```
  ktrem_load ~ normal(load_mu,load_sigma);


  right_othl ~ normal(0,u_sd);
  left_othl  ~ normal(0,u_sd);
  ax_othl    ~ normal(0,u_sd);
  rt_othl    ~ normal(0,u_sd);
  rig_othl   ~ normal(0,u_sd);
  akin_othl  ~ normal(0,u_sd);
  lakin_othl ~ normal(0,u_sd);
  ktrem_othl ~ normal(0,u_sd);

  pc_u_Lcorr ~ lkj_corr_cholesky(20);
  pp_u_Lcorr ~ lkj_corr_cholesky(20);

  for(n in 1:N_pc) {
    pc_eta_raw[n]   ~ std_normal();
    pc_ystar_raw[n] ~ std_normal();
  }

  for(n in 1:N_pp) {
    pp_eta_raw[n]   ~ std_normal();
    pp_ystar_raw[n] ~ std_normal();
  }

  thresh_mu    ~ normal(0,5);
  thresh_sigma ~ cauchy(0,5);

  for(j in 1:J) {
    thresholds[j] ~ normal(thresh_mu,thresh_sigma);
    (thresholds[j] - pp_thresholds[j]) ~ normal(0,thr_diff);
  }

  for(m in 1:M)
    y_pc[m] ~ ordered_logistic(pc_ystar[ii[m],jj[m]],thresholds[jj[m]]);

  for(n in 1:N_pp)
    y_pp[n] ~ ordered_logistic(pp_ystar[n],pp_thresholds);
}

generated quantities {
  matrix[N_pp,J] pp_ll_tmp;
  vector[M] pc_ll_tmp;
  vector[M + (N_pp*J)] log_lik;
  vector[(T-1)] thresh_diff[J];
  real deviance;

  for(j in 1:J)
    thresh_diff[j] = thresholds[j] - pp_thresholds[j];

  for(m in 1:M)
    pc_ll_tmp[m] = ordered_logistic_lpmf(y_pc[m]| pc_ystar[ii[m],jj[m]],
                                                  thresholds[jj[m]]);

  for(n in 1:N_pp)
    for(j in 1:J)
      pp_ll_tmp[n,j] = ordered_logistic_lpmf(y_pp[n,j]| pp_ystar[n,j],
                                                        pp_thresholds[j]);

  log_lik = append_row(pc_ll_tmp,to_vector(pp_ll_tmp));
  deviance = -2.0 * sum(log_lik);
}
```

11.10 Supplementary Material K – Standardised Factor Loadings Table

Item	Right	Left	Axial	Rest Tremor	Rigidity	Akinesia	Kinetic Tremor	Lower Akinesia
3.1	-0.003	-0.007	0.911	0	-0.024	-0.004	0.007	-0.011
3.2	0.022	0.016	0.833	0.008	-0.005	0.005	0.006	0.034
3.3A	0.018	0.083	0.084	0.046	0.538	0.096	0.049	0.059
3.3B	0.384	0.004	0.025	-0.006	0.762	0.023	0.015	0.01
3.3C	-0.008	0.474	0.008	-0.011	0.756	0.004	-0.004	0.001
3.3D	0.113	-0.009	-0.002	-0.003	0.966	-0.01	-0.002	-0.008
3.3E	0.002	0.277	-0.01	0.005	0.915	-0.003	-0.002	0.001
3.4A	0.326	0.019	0.019	0.018	0.027	0.79	0.002	0.014
3.4B	0.003	0.615	0.011	0.001	0.006	0.695	-0.007	0.009
3.5A	0.334	-0.008	-0.002	0.001	-0.01	0.863	0.005	-0.008
3.5B	-0.007	0.625	-0.009	-0.007	-0.01	0.735	-0.001	-0.005
3.6A	0.359	-0.007	-0.004	-0.01	0.009	0.836	0.005	-0.004
3.6B	0.005	0.629	0.009	0.003	-0.005	0.685	0.008	0.007
3.7A	0.478	0.002	0.004	-0.006	0.016	0.011	0.731	0.006
3.7B	0.008	0.523	0.01	-0.007	0.008	0.021	0.656	0.004
3.8A	0.259	-0.002	0.003	-0.001	-0.007	0.002	0.9	-0.002
3.8B	-0.002	0.42	0	0.006	0.001	-0.006	0.828	-0.001
3.9	0.001	0.004	0.899	0.022	-0.001	0.001	0.003	0.002
3.10	-0.005	-0.01	0.936	-0.029	0.027	-0.013	0	-0.005
3.11	-0.017	-0.041	0.455	-0.045	-0.002	-0.025	-0.005	-0.038
3.12	-0.004	0.005	0.833	-0.005	-0.022	0.008	0.023	-0.012
3.13	-0.002	-0.03	0.818	-0.026	0.069	-0.013	-0.032	-0.003
3.14	0.003	0.013	0.93	0.018	-0.009	0.022	0.006	0.008
3.15A	0.617	0.003	0.003	0.011	-0.002	-0.003	-0.001	0.467
3.15B	0.012	0.562	-0.011	0.018	0.003	-0.005	-0.004	0.553
3.16A	0.387	0.001	0.009	-0.008	0.001	0.011	0.002	0.669
3.16B	-0.001	0.453	0.003	-0.015	0.005	0.001	0.006	0.66
3.17A	0.71	-0.004	-0.008	0.412	-0.006	-0.005	-0.007	0.004
3.17B	-0.015	0.496	-0.003	0.664	0.001	-0.005	-0.002	-0.006
3.17C	0.595	-0.002	-0.008	0.422	-0.01	-0.005	0.01	-0.001
3.17D	-0.007	0.419	-0.006	0.579	0.004	-0.015	0.011	-0.01
3.17E	0.005	-0.012	0.076	0.727	0.041	0.055	0.04	0.006
3.18	0.006	0.003	0	0.979	-0.001	0	-0.001	0.004

11.11 Supplementary Material L – Individual Symptom Score Calculator

Please see attached Excel spreadsheet