UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Psychology

Fatigue and salivary cortisol in relapsing-remitting multiple sclerosis: an investigation in everyday life

by

Daniel Powell

Thesis for the degree of Doctor of Philosophy

April 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Psychology

Thesis for the degree of Doctor of Philosophy

FATIGUE AND SALIVARY CORTISOL IN RELAPSING-REMITTING MULTIPLE SCLEROSIS: AN INVESTIGATION IN EVERYDAY LIFE

Daniel Powell

Multiple sclerosis (MS) is a chronic autoimmune disease characterised by inflammatory attacks on the central nervous system. Fatigue is a common symptom in MS, yet its aetiology, exacerbating factors, and manifestation in everyday life are unclear. Given its role in regulating inflammation, energy metabolism and stress responses, cortisol is a relevant psychobiological target for MS fatigue research. The primary aims of this thesis were to examine diurnal fatigue patterns and contextual effects of daily stressors and mood in people with relapsing-remitting MS (RRMS); explore cortisol secretory activity in RRMS in everyday life; and investigate associations of cortisol with fatigue at baseline and change in fatigue 6 months later. Data were collected in an ecological momentary assessment study incorporating repeated real-time self-reports and salivary cortisol assessments over 4 consecutive weekdays in an RRMS group (n = 38) and healthy control group (n = 38), matched for age and sex. Statistical analysis was predominantly by multilevel modelling. The analysis presented in the first empirical chapter (Chapter 3) demonstrated that RRMS fatigue typically follows an increasing (but decelerating) within-day fatigue trajectory, distinct from linear trajectories in controls. Fatigue was sensitive to stressor and mood fluctuations within-subjects in both groups. The analysis presented in the second empirical chapter (Chapter 4) described larger cortisol awakening responses in RRMS compared to controls, but similar diurnal cortisol slopes. Cortisol reactivity to daily life stressors was apparent in both groups, mediated by self-reported distress. A systematic review was conducted (presented in Chapter 5) showing attenuated diurnal cortisol variability is most-frequently associated with fatigue across clinical populations; cortisol output appeared less relevant. However, the analysis presented in the final empirical chapter (Chapter 6) showed greater cortisol variability in the morning (larger cortisol responses to awakening) was associated with RRMS fatigue. Flatter diurnal cortisol slopes were associated with fatigue in controls. Chronic stress and depressive symptoms did not moderate associations, and there was no relationship between cortisol and change in fatigue 6 months later in either group. The original research presented confirms MS fatigue is not a stable symptom experience in everyday life, is sensitive to psychosocial contextual factors, and is associated with cortisol, potentially via pro-inflammatory immune mediators.

Contents

List of Tables

List of Figures

Declaration of Authorship

I, *Daniel Powell*, declare that the thesis entitled *Fatigue and salivary cortisol in relapsing remitting multiple sclerosis: an investigation in everyday life* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Parts of this work have been published as:

Powell, D. J. H., Liossi, C., Moss-Morris, R., & Schlotz, W. (2013). Unstimulated cortisol secretory activity in everyday life and its relationship with fatigue and chronic fatigue syndrome: A systematic review and subset meta-analysis. *Psychoneuroendocrinology, 38(11),* 2405-2422.

Signed: ……..

Date:……

Acknowledgements

The journey to PhD completion has been intellectually and emotionally challenging, and numerous people have helped, supported and provided much-needed perspective throughout. I would like to give special thanks to the supervisory team who have gotten me this far. First and foremost, Dr Wolff Schlotz, who has been a fantastic source of support, advice, enthusiasm and humour, right through from my MSc to the end of my PhD. I am also very grateful to Dr Christina Liossi, who became my primary supervisor at the University along the way, for the detailed feedback and many helpful discussions. Thanks also to Professor Rona Moss-Morris for all the assistance and experience offered throughout, and to Dr Sarah Kirby for being a great advisor always with an open door.

I would like to give special mention to Dr Matthew Jones, who designed the handheld software and provided assistance with programming and empathy when handhelds occasionally misbehaved. Thanks must also go to Carrie Day, Lesley Kempson, and Dr Victoria Williams for their determined support with participant recruitment, and to the Institute of Psychiatry for providing me with a temporary base in London. I also acknowledge the contribution of Chelsea Lane, a Kings College London undergraduate student, who did a fantastic job assisting me with participants while in London. The people with MS who participated have been wonderful to work with, and have taught me so much even from casual conversations; I thank them all for their time, effort, and enthusiasm in supporting the project.

I have been extremely lucky to meet and work with an amazing group of PhD students in Psychology and Health Psychology at the University of Southampton since I began; particularly Jenny Mc Sharry, Ingrid Muller, Gülcan Garip, Sarah Tonkin-Crine, Laura Dennison, and Kat Bradbury, who have all been absolutely brilliant to me from the beginning. Not least, I owe a great deal to my parents, friends and family, who have given me a perfect grounding and support to pursue my interests.

The research in this thesis was funded by an Economic and Social Research Council Studentship and the Psychology Unit at the University of Southampton.

Definitions and Abbreviations

ACTH	Adrenocorticotropic Hormone
AUC	Area Under the Curve
AUCg	Area Under the Curve with respect to Ground
AUCi	Area Under the Curve with respect to Increase
BDI	Beck Depression Inventory
CAR	Cortisol Awakening Response
CBT	Cognitive Behavioural Therapy
CDC	Centre for Disease Control and Prevention
CES-D	Centre for Epidemiologic Studies Depression Scale
CFS	Chronic Fatigue Syndrome
CI	Confidence Interval
CNS	Central Nervous System
CRH	Corticotrophin-Releasing-Hormone
CSA	Cortisol Secretory Activity
CSSS	Chronic Stress Screening Scale
DCS	Diurnal Cortisol Slope
Dex/CRH	Combined Dexamethasone Corticotrophin-Releasing-Hormone
DMT	Disease Modifying Therapy
DST	Dexamethasone Suppression Test
EAE	Experimental Autoimmune Encephalomyelitis
EDSS	Expanded Disability Status Scale
EMA	Ecological Momentary Assessment
FS	Fatigue Scale
FSN	Fail Safe N
FSS	Fatigue Severity Scale
Gd+	Gadolinium-Enhancing

HADS	Hospital Anxiety and Depression Scale
HPA	Hypothalamic-Pituitary-Adrenal
IDA	Iron Deficiency Anaemia
IFN	Interferon
IL	Interleukin
KMO	Kaiser-Meyer-Olkin
MESOR	Midline Estimate Statistic Of Rhythm
MF	Momentary Fatigue
MFIS	Modified Fatigue Impact Scale
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MUS	Medically Unexplained Symptoms
NICE	National Institute for Health and Care Excellence
PDDS	Patient Determined Disease Steps
PIS	Participant Information Sheet
PPMS	Primary Progressive Multiple Sclerosis
PSS	Perceived Stress Scale
pwMS	People with Multiple Sclerosis
pwRRMS	People with Relapsing-Remitting Multiple Sclerosis
RCT	Randomised Controlled Trial
RRMS	Relapsing-Remitting Multiple Sclerosis
S1	Sample 1
SD	Standard Deviation
SPMS	Secondary Progressive Multiple Sclerosis
TNF	Tumor Necrosis Factor
T*x*	Sample provided *x* minutes after awakening
UV	Ultraviolet
VAS	Visual Analogue Scale

Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease characterised by inflammatory demyelination of the central nervous system (Compston & Coles, 2008; Noseworthy, Lucchinetti, Rodriguez, & Weinshenker, 2000). There are approximately 127,000 people with MS (pwMS) in the UK (Mackenzie, Morant, Bloomfield, MacDonald, & O'Riordan, 2014). MS aetiology is uncertain and no cure exists, meaning clinical efforts are limited to the management of symptoms and slowing down of neurological damage and functional loss. Fatigue is commonly described by pwMS as being among their worst and most-disabling symptoms, and is considered sensitive to daily demands (Fisk, Pontefract, Ritvo, Archibald, & Murray, 1994; Mills & Young, 2008); however, the aetiology of MS fatigue is unclear with a number of disease-mediated and secondary mechanisms proposed (Induruwa, Constantinescu, & Gran, 2012; Kos, Kerckhofs, Nagels, D'hooghe, & Ilsbroukx, 2008; Krupp, Serafin, & Christodoulou, 2010). Cortisol is an important endocrine regulator of inflammation, energy metabolism and the stress response (Sapolsky, Romero, & Munck, 2000), and therefore has several functions making it a relevant psychobiological target for MS fatigue research. This doctoral thesis presents a prospective examination of fatigue and cortisol from an everyday life perspective in a case-control ecological momentary assessment (EMA) study. In this brief introduction, a short rationale is provided for this doctoral research. A more critical evaluation of previous research is given in the literature review presented in Chapter 1 and in respective empirical chapters.

MS is a heterogeneous disease categorised into three main types: relapsing-remitting MS (RRMS); secondary progressive MS (SPMS); and primary progressive MS (PPMS). The clinical group of interest in this thesis is RRMS: the most-common MS-type. In 80-85% of cases, initial MS diagnosis is RRMS, which is characterised by episodic symptom exacerbations lasting 1-4 weeks (relapse-phase), followed by complete or partial recovery to previous levels of functioning (remission-phase). Relapses are unpredictable and vary in frequency between individuals, but there is accumulating evidence that risk of relapse in RRMS is increased by stressful life events (Artemiadis, Anagnostouli, & Alexopoulos, 2011; Mohr, Hart, Julian, Cox, & Pelletier, 2004). RRMS changes to a progressive disease course (SPMS) within 15 years in approximately 65% of cases, where symptoms gradually worsen over time, with or without acute

relapses (Koch, Uyttenboogaart, van Harten, & De Keyser, 2008). Initial diagnosis in the remaining 15-20% is PPMS, where there is progressive decline in functioning from onset, without relapses (D. H. Miller & Leary, 2007). People with RRMS tend to accrue disability slower than individuals with progressive MS-types (Tremlett, Zhao, Rieckmann, & Hutchinson, 2010) and are clinically-stable when in remission-phase. These characteristics contributed to the rationale for recruiting people with RRMS (in remission-phase) only; participants required a good level of fine motor skills and cognitive functioning to carry out the EMA protocol reliably in everyday life. A more comprehensive overview of eligibility criteria is presented in Chapter 2 as part of a critical discussion of the investigation protocol.

Large-scale (n >500) studies have suggested 60-85% of pwMS experience symptomatic fatigue (Hadjimichael, Vollmer, & Oleen-Burkey, 2008; Lerdal, Celius, & Moum, 2003; Minden et al., 2006). The level of fatigue in RRMS is typically similar to fatigue in progressive MS-types, after adjusting for neurological disability (Kroencke, Lynch, & Denney, 2000). MS fatigue is considered to have a negative impact on daily living (Flensner, Ek, & Söderhamn, 2003) and is associated with lower quality of life (Amato et al., 2001). Importantly, fatigue severity is frequently described by pwMS as fluctuating rather than stable and consistent (Freal, Kraft, & Coryell, 1984; Mills & Young, 2008). However, current understanding of MS fatigue aetiology is based primarily on cross-sectional research using retrospective self-reports, which eliminate all within-subjects variability, lack ecological validity, and may also be affected by recall bias (Hufford, 2008; Reis, 2011; N. Schwartz, 2011). Operationalising MS fatigue using summary measures discards important information about symptom experience. Several qualitative explorations have reported pwMS generally consider fatigue to be exacerbated by psychosocial stress (Mills & Young, 2008; Stuifbergen & Rogers, 1997). The effect of stress on fatigue in MS has never been prospectively tested in everyday life. Using an EMA strategy permits the repeated measurement of phenomena in real-time and in real-world settings (with high ecological validity), and can enable the examination of within-subjects contextual effects.

The hypothalamic-pituitary-adrenal (HPA) axis is a self-regulatory endocrine cascade pathway that secretes its end-product, cortisol, in a circadian rhythm, and is sensitive to both psychological and physiological

stress (Dickerson & Kemeny, 2004; D. B. Miller & O'Callaghan, 2002). Cortisol is relevant to fatigue due to its regulatory role in energy metabolism, the high levels of fatigue present in conditions characterised by low cortisol (e.g., Addison's disease), and the relative efficacy of low-dose corticosteroid treatments in alleviating fatigue (for short periods) in chronic fatigue syndrome (Cleare, 2003; Sapolsky et al., 2000; Ten, New, & Maclaren, 2001). Reduced cortisol outputs are commonly described in chronic fatigue syndrome (Cleare, 2003; Papadopoulos & Cleare, 2012); however, whether HPA axis dysfunction is also relevant to fatigue experienced in other clinical populations is less clear. Cortisol can be measured in saliva, blood, urine, and hair, but salivary cortisol is the most popular method for studies of HPA axis activity in everyday life due to sampling being non-invasive and easy for participants to self-administer (Kirschbaum & Hellhammer, 1989).

In recent years, cortisol secretory activity (CSA) in everyday life in MS has been examined in studies by 2-day salivary cortisol protocols (Gold et al., 2010; Gold et al., 2011; Kern et al., 2013; Kern et al., 2011). Typically, these studies have reported facets of CSA are elevated in pwMS (specifically, RRMS) compared to healthy individuals. However, there were methodological concerns with each of these studies; a critical overview will be provided in upcoming chapters. The salivary cortisol study by Gold et al. (2011) presents the only test of associations between facets of unstimulated basal CSA and fatigue in MS, with no statistically significant association found (*r* not reported). However, this association was tested in an RRMS sample where 23% of participants also had comorbid major depressive disorder: a psychiatric condition where fatigue is an important somatic symptom (American Psychiatric Association, 2000) and hyperactivity of the HPA axis is frequently reported (Pariante & Lightman, 2008). Therefore, the relevance of cortisol to MS fatigue remains uncertain.

The primary aims of the research presented in this thesis were to provide a detailed investigation of (1) fatigue experience within-individuals and within-days; (2) the circadian rhythm of cortisol in RRMS; and (3) associations between RRMS fatigue and cortisol both at baseline and longitudinally. The thesis begins with a literature review in Chapter 1 providing a detailed overview of the nature of MS and MS fatigue, before discussing the HPA axis and its potential relevance to fatigue in MS. Chapter 2 begins with a critical discussion of the rationale behind EMA, before presenting the general investigation

protocol for this doctoral research: a case-control study including 4 consecutive weekdays of EMA with salivary cortisol assessments and momentary self-reports at strategic time-points. As the data produced by EMA was hierarchical (assessments nested within days, nested within participants), multilevel modelling was utilised as the most appropriate statistical analysis method (Hox, 2010; Singer & Willett, 2003; Snijders & Bosker, 2012). Chapter 2 concludes by providing a general overview of the advantages to using multilevel modelling for hierarchical EMA data.

Chapter 3 presents the first original analysis of the EMA data, examining the experience of fatigue in RRMS in everyday life, compared to fatigue experienced by healthy individuals. The chapter focuses on typical diurnal fatigue patterns (changes in fatigue over time within days) and the contextual effects of psychosocial stressors and momentary mood on fatigue severity within individuals. In a supplementary analysis, Chapter 3 also investigates the accuracy of end-of-day self-reports of recalled fatigue in terms of discrepancy with aggregated momentary fatigue ratings and presence of peak (most-severe fatigue) and end (most-recent fatigue) cognitive biases. As pwMS are typically more familiar with the experience of fatigue - and the reporting of fatigue - than healthy individuals, it was expected that the RRMS group would provide more-accurate self-reports for recalled fatigue.

Chapter 4 presents an analysis focussing on group comparisons for salivary cortisol outcomes, to investigate HPA axis activity in everyday life in RRMS. The cortisol awakening response (CAR) and diurnal cortisol slope were both computed as outcomes from constituent saliva samples. These markers of HPA axis activity will be explained in more detail in upcoming chapters. In addition, a within-subjects analysis is presented examining the functioning of the HPA axis-related stress response in RRMS by investigating group differences in the cortisol response to daily life stressors. Chapter 5 then presents a systematic review investigating the relationship between cortisol and fatigue across all populations, including chronic fatigue syndrome and other clinical and non-clinical populations. The review focussed solely on studies investigating CSA using salivary cortisol assessments in everyday life. The aim of this review was to explore whether certain facets of CSA were more or less relevant to fatigue, and whether associations were similar in all populations where fatigue is experienced.

Chapter 6 presents the final analysis of the EMA data, examining cortisol-fatigue associations in both groups, and incorporating cortisol measured at baseline and fatigue measured at baseline and at 6-month follow-up. This analysis represents the first longitudinal examination of cortisol-fatigue associations in any clinical population. The CAR is often considered to prepare the individual for the demands of the upcoming day, although the exact mechanism is unknown (Fries, Dettenborn, & Kirschbaum, 2009; Schlotz, Hellhammer, Schulz, & Stone, 2004; Schulz, Kirschbaum, Pruessner, & Hellhammer, 1998). A within-subjects analysis presented in Chapter 6 extends this hypothesis to examine whether the CAR predicts upcoming-day fatigue in RRMS.

Lastly, a general discussion is presented in Chapter 7 which provides a review of the results in this thesis and evaluates the generalisability of the findings by critically examining the representativeness of the recruited sample. Overall implications of these findings are discussed, along with a discussion of the limitations and suggestions for future research.

Chapter 1: Multiple sclerosis, fatigue, and cortisol

This literature review begins by presenting the key features of MS, including its diagnostic criteria, proposed aetiology, symptoms, treatments, and prognosis. Current literature around the definition, measurement, and aetiology of MS fatigue is then critically examined. An introduction to the hypothalamic-pituitary-adrenal (HPA) axis is then provided, with rationale for examining its product, cortisol, as relevant to both MS and, specifically, MS fatigue. Current research examining the HPA axis in MS is then critically reviewed. The chapter concludes by presenting the research aims for this thesis. To assist with interpreting the literature presented, Cohen's *d* effect sizes (J. Cohen, 1988) are presented with 95% confidence intervals (CI) where possible. Where studies have not reported effect sizes but present sufficient data, Cohen's *d* effect sizes were derived manually using formulae based on independent or paired samples t-tests (see Appendix A). This procedure is consistent throughout the thesis. Presented data indicate mean (*M*) ± standard deviation (*SD*).

1.1 The nature of multiple sclerosis

MS is a chronic autoimmune disease characterised by inflammatory demyelination of the CNS (Noseworthy et al., 2000), and is among the most common neurological disorders with currently around 127,000 people with MS (pwMS) in the UK (Mackenzie et al., 2014). Despite extensive research, the cause of MS is unknown and no proven cure exists (Compston & Coles, 2008). Onset typically occurs between 20 and 40 years of age (Richards, Sampson, Beard, & Tappenden, 2002), and women are diagnosed two-to-three times as often as men (Alonso & Hernán, 2008; Noonan, Kathman, & White, 2002); recent UK data suggested the prevalence of MS was 285.8 per 100,000 women and 113.1 per 100,000 men (Mackenzie et al., 2014). MS prevalence increases with distance from the equator (Alonso & Hernán, 2008; Simpson, Blizzard, Otahal, Van der Mei, & Taylor, 2011) and is at its highest in white populations and in high-income countries (Koch-Henriksen & Sørensen, 2010).

1.1.1 Pathophysiology and aetiology

MS is believed to be mediated by autoimmune processes, with inflammatory attacks on the myelin sheath surrounding nerve fibres (demyelination) in the CNS by the body's own immune system (Korn, 2008; McFarland & Martin, 2007). The myelin sheath serves to protect and stabilise nerve axons, and facilitates the conduction of electrical nerve impulses. Loss of myelin results in the formation of inflammatory lesions or plaques (sclerosis), disrupting usual CNS communication and normal functioning. Lesions represent a breakdown of the blood-brain barrier and presence of inflammation. Neural damage is thought to be the key contributor to long-term disability in MS (Chaudhuri & Behan, 2005).

1.1.1.1 Diagnostic criteria and MS-types

An MS diagnosis is given by a neurologist using the accepted "McDonald criteria" (McDonald et al., 2001) which have since been revised (Polman et al., 2011; Polman et al., 2005). The criteria stipulate an MS diagnosis requires at least two distinct episodes of disease activity (lesions or clinical signs) occurring at different times and affecting at least two of four CNS regions: periventricular, juxtacortical, infratentorial, or spinal cord.

Episodes can be confirmed by incorporating clinical history assessments and magnetic resonance imaging (MRI). MRI plays a crucial role in MS diagnosis, facilitating the examination of lesion formulation using gadolinium-enhancing (Gd+) MRI. The gadolinium contrast agent is injected into the bloodstream and enters the brain at lesions via blood-brain barrier disruption and appears vividly on MRI (Neema, Caccarelli, Jackson, & Bakshi, 2012). New lesions are enhanced by gadolinium for 3-4 weeks while older lesions appear as darker spots (Neema et al., 2012). Therefore, previous episodes of disease activity can be detected on MRI even where no symptoms presented.

The pattern of disease activity informs the classification of an individual's MS into one of three categories: relapsing-remitting MS (RRMS); secondary progressive MS (SPMS); or primary progressive MS (PPMS). Around 80-85% of pwMS are diagnosed with RRMS at onset, where the disease course is characterised by episodic relapses (symptom exacerbations) lasting 1-4 weeks before partial or full remission to previous levels of functioning for months or sometimes years (Noseworthy et al., 2000; Tremlett et al., 2010). A relapse will

present a new or returning symptom, and relapse rates vary markedly between individuals. A longitudinal study of 2477 pwMS reported women had a 14.3%, 95% CI [5.8, 23.3], higher relapse-rate than men, and relapse-rate decreased by 17% for every 5 years post-onset (Tremlett, Zhao, Joseph, & Devonshire, 2008). In around 65% of cases, RRMS evolves into SPMS within 15 years of onset (Koch et al., 2008). SPMS is not an initial diagnosis; it cannot be diagnosed at onset. SPMS entails periods of steady neurological and symptom deterioration between relapses (Richards et al., 2002) meaning disability accumulates more quickly than with RRMS previously.

The remaining 15-20% of initial diagnoses are of PPMS, with steady accrual of neurological damage over time from onset, rarely with relapses (D. H. Miller & Leary, 2007). A PPMS diagnosis is permitted where an individual presents with a clinically progressive course lasting more than 12 months (prospectively or retrospectively determined) with evidence of at least two of the following: (1) one or more lesions in at least one of the periventricular, juxtacortical, or infratentorial areas; (2) two or more lesions in the spinal cord; (3) evidence of oligoclonal bands (proteins suggestive of CNS inflammation) in cerebrospinal fluid examined by lumbar puncture (Polman et al., 2011). Compared with RRMS, PPMS tends to be diagnosed at an older age (on average, 40 years old at diagnosis) and the gender ratio is more equal (approximately 1:1) (D. H. Miller & Leary, 2007). Prognosis is worse in PPMS; in a review of longitudinal studies, the median times reported from onset to requiring a cane to walk was 15-32 years in RRMS-onset and less than 10 years in PPMS-onset (Tremlett et al., 2010). In 10-20% of all pwMS, a retrospective diagnosis of benign MS is given when little or no symptom deterioration has occurred over a prolonged period of time, typically 15 years (Amato et al., 2006).

1.1.1.2 Genetic and environmental factors

Although MS aetiology is uncertain, it is likely influenced by multiple factors, including genetics and the environment. An epidemiological study found that Human Leukocyte Antigen type frequency, ultraviolet (UV) radiation index, and cigarette smoking accounted for approximately 75% of the variance in MS prevalence rates across Europe (Handel, Handunnetthi, Giovannoni, Ebers, & Ramagopalan, 2010). Human Leukocyte Antigen alleles reflect different types of cell surface molecules (major histocompatibility complex)

encoded by genes, and mediate immune activity at the cellular level: a potential mechanism for genetic involvement in MS pathology.

Twin studies have explored genetic precursors to MS. Willer et al. (2003) found a concordance rate of 25-30% in monozygotic (identical) twins versus approximately 5% in dizygotic (non-identical) twins and 3% in non-twin siblings. Therefore, while genetic factors seem to influence susceptibility to MS, additional factors are likely involved. Baranzini et al. (2010) were unable to identify any genomic or epigenetic (genetic expression) differences in immune cells of monozygotic twins where only one had developed MS, indicating no genetic mutations had occurred to trigger disease onset. These findings suggest a considerable environmental influence over acquiring MS, but individuals may retain genetic predispositions requiring an environmental "cue" (trigger).

High levels of UV radiation were associated with less MS prevalence in the study by Handel et al. (2010). Assuming UV radiation is relevant, vitamin D levels may elucidate the mechanism underlying a protective role; UV exposure is known to stimulate cutaneous (skin-related) vitamin D synthesis (Gilchrest, 2008). In a longitudinal study of over 180,000 women in the US (Munger et al., 2004), those taking vitamin D supplements (≥400 international units/day) had a relative risk of developing MS, compared to those who did not take supplements, of 0.59, 95% CI [0.38, 0.91], suggesting a protective effect of vitamin D. However, this association may be confounded by the high correlation of UV/vitamin D with latitude, and may therefore be an artefact of another factor.

The "hygiene hypothesis" proposes that minimal exposure to infectious agents in western societies, particularly in infancy, provides restricted opportunity to develop an efficiently-regulated immune system, predisposing individuals to allergic or autoimmune reactions later in life (Strachan, 1989, 2000). For example, MS incidence was minimal (<20 per 100,000) in countries with >10% prevalence of *Trichuris trichiura* (a parasitic worm) infection, but far higher (120 per 100,000) in countries with <10% prevalence (Fleming & Cook, 2006). The hygiene hypothesis proposes multiple microorganisms are involved, rather than suggesting a specific pathogen as prerequisite to MS. This non-specificity has been criticised, with observations in animal models of autoimmune diseases demonstrating very different immune responses to

specific infectious agents (Bach, 2002). As such, it is unlikely that all agents have an equal potential to protect against autoimmune diseases such as MS (Ascherio & Munger, 2007). Rather than be protective, Epstein-Barr virus has been proposed as a possible risk factor for MS, with evidence of elevated Epstein-Barr virus antibodies prior to onset (Ascherio & Munger, 2010); however, causality cannot yet be inferred (Pakpoor, Giovannoni, & Ramagopalan, 2013).

1.1.1.2.1 Involvement of stress in MS onset and risk of relapse

A causative role for stress in MS onset is a (relatively) commonly-held belief among pwMS. A survey of 2529 pwMS found 27% felt "stressful or traumatic events" caused their MS: the most prevalent causal attribution made (Simmons, Ponsonby, van der Mei, & Sheridan, 2004). However, evidence implicating stress in MS onset is largely limited to studies of the predominant animal (rat/mice) model of MS: experimental autoimmune encephalomyelitis (EAE). A review described differential effects on EAE onset by chronic (repeated or long-lasting) and acute (short-term) stressors; chronic stress tended to delay EAE onset and suppress disease processes, whereas acute stress usually enhanced them (Heesen, Gold, Huitinga, & Reul, 2007). It is debatable whether findings from EAE studies can be generalised to MS. MS appears an exclusively human disease, with no other animal spontaneously developing a similar condition (Baker, Gerritsen, Rundle, & Amor, 2011); many pathological differences exist between EAE and MS, including lesions in EAE located predominantly in lumbar regions, which is rare in MS (Sriram & Steiner, 2005). Although generalising findings in EAE to MS is controversial, several pharmaceutical treatments for MS have evolved from experimentation in EAE models (Steinman & Zamvil, 2006).

In humans, evidence associating stress with MS onset is limited. From the few studies published, a systematic review concluded there was evidence to suggest stress can increase risk of MS onset (Artemiadis et al., 2011). However, conclusions were drawn from studies retrospectively quantifying stress over a period of time in individuals recently diagnosed with MS compared to healthy controls. For example, in 41 newly-diagnosed pwMS and 41 healthy controls, the MS group reported significantly more negative life events, d = 0.64, 95% CI [0.19, 1.07], and problems in family life, d = 0.57, 95% CI [0.12, 1.00], in the prior 3 years than controls (X. J. Liu et al., 2009). In addition, Artemiadis et al.

(2011) concluded longitudinal studies suggested stress was a risk factor for relapses in RRMS, but highlighted heterogeneity in stress measurement as a reason for interpreting findings with caution. An earlier meta-analysis also concluded stressful life events were associated with increased relapse risk in RRMS, d = 0.53, 95% CI [0.40, 0.65] (Mohr et al., 2004) although perceptions of stress in constituent studies may be biased by simultaneously elevating disease activity affecting appraisals, rather than a real difference in stressful experience. Stress-management interventions have since shown reductions in Gd+ lesion frequency in intervention groups (Mohr et al., 2012) and a 46.1% reduction in MS symptom-intensity generally (Artemiadis et al., 2012).

1.1.2 Symptoms and disability

With demyelination affecting any CNS region, wide-ranging symptoms can present. Fatigue is frequently reported as the most common symptom in MS (e.g., Freal et al., 1984; Minden et al., 2006). A study surveying 2156 pwMS reported currently experienced symptoms included fatigue (83%), difficulty walking (67%), stiffness and spasms (63%), cognitive problems (56%), bladder (60%) and bowel (35%) dysfunction, pain (54%), and several others (Minden et al., 2006). These rates correspond well with other reports (e.g., Stazzone & Brown, 2012).

Very few symptoms are MS-specific. Lhermitte's symptom, describing an "electrical sensation running down the spine or limbs on neck flexion" is considered a classical sign of MS (Compston & Coles, 2008, p. 1502), and is usually a recurring transient symptom. However, Lhermitte's is not experienced by all, with around 33-40% experiencing Lhermitte's at some stage since MS diagnosis (Al-Araji & Oger, 2005; Kanchandani & Howe, 1982). Uhthoff's phenomenon is common and describes a temporary worsening of symptoms when core body temperature rises (Compston & Coles, 2008).

A recent study of 27,918 pwMS providing longitudinal self-reports of disability from 1996 onwards produced a reference table for patients and clinicians outlining typical disability levels after certain MS durations (Kister et al., 2013). For example, within the first year of diagnosis, 44% reported no or mild symptoms with no limitations on daily activities (Patient-Determined Disease Steps score = 0), 85% did not yet need a cane for walking (PDDS < 4) and 1% required a wheelchair (PDDS = 7). By 10 years post-diagnosis, 19%

scored 0 on the PDSS, 60% scored less than 4, and 8% scored 7; and by 20 years post-diagnosis, 10% scored 0, 44% scored less than 4, and 16% scored 7. Although these findings are useful, the results did not consider MS-type, which was surprising given that disability progression occurs more rapidly in PPMS and SPMS than RRMS.

1.1.3 Treatment

With no cure imminent, management of relapses and symptoms, and slowing disease progression are the main treatment goals in MS. Acute relapses, characterised by enhanced neural inflammation and disability, are treated with high-dose methylprednisolone (a synthetic corticosteroid) administered either intravenously (500–1000mg daily for 3-5 days) or orally (500-2000mg daily for 3-5 days) (NICE, 2003). Methylprednisolone is thought to shorten relapse duration and severity by reducing inflammation. Intravenous high-dose methylprednisolone regimens (500mg daily for 5 days) significantly reduced disability scores at 1-week and 4-weeks post-administration compared to a placebo condition (saline) in a randomised controlled trial (RCT) (Milligan, Newcombe, & Compston, 1987). A recent systematic review concluded that oral and intravenous methylprednisolone were equally efficacious in treating MS relapses (Burton, O'Connor, Hohol, & Beyene, 2012).

Disease modifying therapies (DMT) aim to reduce relapse frequency and associated disability progression in RRMS and SPMS. The most common DMTs include daily injections of interferon-beta 1α, interferon-beta 1β, or glatiramer acetate, with studies demonstrating reduced relapse-rates (around 30%) and disability progression (40%) over 2-3 years (Comi, Filippi, & Wolinsky, 2001; Kappos et al., 2007; PRISMS Study Group, 1998). DMTs do not reverse deficits already acquired, and are only effective in SPMS where there are still frequent relapses (Compston & Coles, 2008). Natalizumab is a DMT administered intravenously 4-weekly in outpatient clinics, and has been shown to reduce relapse risk (40%) and disease progression (25%) over a 2-year period in RRMS (Pucci et al., 2011). In the UK, natalizumab is only administered in individuals with rapidly-evolving RRMS (more than two relapses per year and more than one Gd+ lesion on MRI) (NICE, 2010). There are currently no DMTs with clinical benefits for PPMS or SPMS without relapses, so treatment is limited to targeting

symptoms themselves, such as oxybutynin for bladder dysfunction and baclofen for pain and spasms (Murray, 2006).

1.1.4 Prognosis

MS onset is not immediately life-threatening; in a longitudinal US cohort study observing 441 pwMS over 11 years, median survival time was 33.0 years from MS diagnosis and 38.0 years from the onset of symptoms (Hirst, Swingler, Compston, Ben-Shlomo, & Robertson, 2008). However, there is evidence that pwMS experience shorter lifespans than the general population. Based on the General Practice Research Database (a large representative UK sample), life expectancy for males with MS was 12.9 years shorter than the UK average (78.3 years) and, for females, 10.2 years shorter than the UK average (81.8 years) (Mackenzie et al., 2014). In another study, median survival time from MS onset was 10 years shorter than in an age-matched general population (Brønnum-Hansen, Koch-Henriksen, & Stenager, 2004). While it is thought around two-thirds of deaths in MS populations can be attributed to MS and the associated risk of infection (particularly skin, chest, and bladder) (Compston & Coles, 2008), it is often difficult to determine the ultimate cause of death as MS. For example, Hirst et al. (2008) found MS was the explicit "cause of death" in death certificates in only 9% of 221 pwMS who died during observation.

1.1.5 Comorbidities

Physical comorbidities were present (by self-report) in 36.7-77.1% of 8983 pwMS (Marrie et al., 2008); the lower estimate was based on "very likely" accurately reported conditions (e.g., cancer and diabetes) and the higher estimate based on all self-reported physical comorbidities. Unsurprisingly, those pwMS aged over 60 years were far more likely to have physical comorbidity than those under 44 years, *OR* = 5.91, 95% CI [4.95, 7.06].

Common psychiatric morbidity in MS has long been noted (Charcot, 1887) and lifetime prevalence of depression in MS is estimated at 50% (Goldman Consensus Group, 2005). A large health survey (*n* = 115,071) revealed annual prevalence rates of major depressive disorder (MDD) were over twice as high in pwMS, 15.7%, 95% CI [10.9, 20.6], than individuals without MS, 7.4%, 95% CI [7.2, 7.6], and higher than individuals with other chronic conditions, 9.1%,

95% CI [8.9, 9.4] (Patten, Beck, Williams, Barbui, & Metz, 2003). Accounting for age and gender, prevalence of MDD in MS was over twice that of people without MS, *OR* = 2.3, 95% CI [1.6, 3.3]. Depressive symptoms are also far higher in MS than healthy individuals according to a meta-analysis, *d* = 1.07, 95% CI [0.86, 1.28] (Dalton & Heinrichs, 2005). Of note, MDD in MS may also be underdiagnosed due to significant symptom overlap, including fatigue (Goldman Consensus Group, 2005).

The factors contributing to the high prevalence of MDD in MS are unclear. While MDD may be the result of enhanced cerebral inflammation, it may equally be a psychological response to the uncertainties and limitations imposed by what is a progressive and disabling disease (Compston & Coles, 2008). Resolution of MDD is usually accompanied by reduced pro-inflammatory cytokine synthesis; pro-inflammatory cytokines can induce sleep disturbance and loss of appetite which may sustain the propensity to feel depressed (Ziemssen, 2009). Arnett, Barwick, and Beeney (2008) proposed a model of depression in MS, which suggested disease-related neurological, immunological, and other physiological changes represent distal risk factors for depression, but do not explain all variance. Further disease-related changes in fatigue, pain, cognitive and physical functioning may also be associated with depression, with the model proposing these associations are moderated by stress, coping, social support, and conceptions of self and illness. While this model is helpful in understanding potential risk factors for MDD in MS, it was primarily based on cross-sectional data and, as yet, limited causal relationships can be elucidated.

1.2 Multiple sclerosis fatigue

Fatigue is a common symptom experienced by pwMS and often described as among the worst symptoms (Fisk, Pontefract, et al., 1994). MS fatigue is usually a chronic symptom, and can be the first symptom experienced weeks or even months before the first attack, as well as precede and accompany acute relapses in RRMS (Comi, Leocani, Rossi, & Colombo, 2001). Early reports listed fatigue as a symptom in 78-96% of pwMS (Freal et al., 1984; Murray, 1985); however, both studies failed to define fatigue for respondents and did not include a control group to examine the proportion of healthy individuals also reporting fatigue. These criticisms were later addressed by Krupp, Alvarez,

LaRocca, and Scheinberg (1988), defining fatigue as “a sense of physical tiredness and lack of energy, distinct from sadness or weakness” and found 88% of pwMS and 51% of healthy individuals stated they were “bothered by fatigue”. Cross-sectional prevalence rates of MS fatigue have typically been 60-85% in large-sample (*n* >500) studies (Hadjimichael et al., 2008; Lerdal et al., 2003; Minden et al., 2006). Response rates in these studies were low (27.4-53.0%), meaning reported fatigue prevalence may not reliably represent its true prevalence in the MS population.

Studies tend to report higher levels of fatigue in progressive MS-types than RRMS (Flachenecker et al., 2002; Kroencke et al., 2000; Mills & Young, 2010; Patrick, Christodoulou, & Krupp, 2009; Pittion-Vouyovitch et al., 2006), although this is likely due to enhanced neurological disability in PPMS and SPMS (Kroencke et al., 2000; Pittion-Vouyovitch et al., 2006). However, in other studies, disability was not associated with fatigue (Lerdal et al., 2003) and the disability-fatigue relationship was weakened (association attenuated and statistical significance lost) once depression was accounted for (Bakshi et al., 2000).

After controlling for disability and depressive symptoms, MS fatigue has been associated with lower quality of life across several domains: health perception, role limitations due to physical dysfunction, and social function (Janardhan & Bakshi, 2002). Fatigue was also independently associated with reduced quality of life elsewhere (Amato et al., 2001). The impact of MS fatigue on daily life includes inhibiting participation with the outside world (Olsson, Lexell, & Söderberg, 2005), and having a negative effect on employment (Johnson et al., 2004), social activities (Malcomson, Lowe-Strong, & Dunwoody, 2008), and inter-personal communication (B. E. Blaney & Lowe-Strong, 2009). MS fatigue is an important issue for healthcare providers; a longitudinal study monitoring health service use over 30 months found pwMS with persistent fatigue made, on average, 156% more primary care visits and 70% more hospital outpatient visits than pwMS without persistent fatigue (Johansson, Ytterberg, Gottberg, Widén Holmqvist, & von Koch, 2009).

1.2.1 Definition and measurement

Fatigue is a complex concept to define and measure, with little consensus achieved. Fatigue in chronic illness is generally agreed as best described as a

multifaceted construct embodying biological, psychological, and behavioural processes (Swain, 2000). Chaudhuri and Behan (2000) described two distinct components of fatigue in neurological disorders: central and peripheral fatigue. Central fatigue is perceived by the CNS (subjective) and most relevant to disorders like MS, whereas peripheral fatigue is localised muscle-weakness (objective) and most relevant to neuromuscular disorders. MS fatigue should not be confused with "weakness" as both fatigue and weakness can occur in the other's absence (Bakshi, 2003; Chaudhuri & Behan, 2000).

Mills and Young (2008) conducted semi-structured interviews with 40 pwMS (19 RRMS, 16 SPMS, 5 PPMS) to present a clinical definition of MS fatigue summarising participants' conceptualisations of the symptom. A variety of different semantics were used by participants to describe fatigue, including "tiredness", "weariness", and being "shattered", "exhausted", or "zonked" (p.52). The study highlighted various fatigue manifestations, including heaviness in limbs, loss of pre-learned motor skills, physical and mental exhaustion, and lack of motivation. The concluding summary definition was comprehensive:

> In summary, [MS] fatigue is defined as reversible, motor and cognitive impairment with reduced motivation and desire to rest, either appearing spontaneously or brought on by mental or physical activity, humidity, acute infection and food ingestion. It is relieved by daytime sleep or rest without sleep. It can occur at any time but is usually worse in the afternoon. In MS, fatigue can be daily, has usually been present for years and has greater severity than any premorbid fatigue. (Mills & Young, 2008, p.57)

The Mills and Young (2008) definition has a distinctly physiological basis with little explicit emphasis on the subjective nature of fatigue. However, the definition does confirm within-day variability in fatigue severity and raises the question, among others, of what underlies changes in fatigue within days, and the types of activities that may induce or exacerbate fatigue (Mills & Young, 2008).

Several other definitions of MS fatigue exist that are concise but seem to misrepresent its complexity, including "difficulty initiating or sustaining voluntary effort" (Chaudhuri & Behan, 2004), "feelings of physical tiredness and

lack of energy distinct from sadness or weakness" (Krupp et al., 1988), and "an overwhelming sense of tiredness, lack of energy or feelings of exhaustion" (Comi, Leocani, et al., 2001). The most commonly-used definition of MS fatigue was provided within the Multiple Sclerosis Council for Clinical Practice Guidelines (1998) as a "subjective lack of physical and/or mental energy that is perceived by the individual or caregiver to interfere with usual and desired activities" (p.2). The MS Council definition clearly defines MS fatigue as subjective, whilst acknowledging physical and mental components as well as its impact on daily living. Recent reviews of MS fatigue have all described the MS Council definition as the predominant definition within the research field (Bol, Duits, Hupperts, Vlaeyen, & Verhey, 2009; Induruwa et al., 2012; Kos et al., 2008; Krupp et al., 2010).

MS fatigue measurement is equally complex, with its subjectivity requiring self-report. Self-report is often criticised for the recall bias present in assessments requiring retrospective summation (Schwid, Covington, Segal, & Goodman, 2002). A variety of self-report tools have been developed to measure recalled fatigue in chronic illness, with varying psychometric qualities (Whitehead, 2009). In a systematic review, 252 unique measures of subjective fatigue were identified (Hjollund, Andersen, & Bech, 2007), including multidimensional measures and others solely focussed on symptom severity or impact. For example, the Fatigue Scale (FS; Chalder et al., 1993) and the Fatigue Severity Scale (FSS; Krupp, LaRocca, Muir-Nash, & Steinberg, 1989) are two commonly-used measures of fatigue severity in MS; whereas the Modified Fatigue Impact Scale (MFIS; Ritvo et al., 1997) is frequently used to measure the impact of fatigue on daily life in MS. Currently, little consensus exists on a "gold standard" method of measuring MS fatigue in research or clinical practice (Flachenecker et al., 2002).

Momentary reports (measured at the point of experience) offer a unique perspective to fatigue measurement: the ability to detect within-individual and within-day variability. Momentary measures minimise self-report recall bias as reconstruction of experience from autobiographic memory is not required (Hufford, 2008). Studies in other populations utilising momentary fatigue measures have demonstrated that fatigue experience is relatively stable for some while an erratic phenomenon for others (Hacker & Ferrans, 2007; Stone, Broderick, Porter, & Kaell, 1997).

1.2.2 Demographic features of MS fatigue

The largest study of the epidemiology of fatigue in MS was carried out within the North American Research Committee on MS Patient Registry, with 9205 of 18595 (49.5%) registrants responding (Hadjimichael et al., 2008). Participants were divided into "severe fatigue" (n = 6691) or "mild/moderate fatigue" (n = 2386) groups based on an FSS score cut off (>36; equivalent of ≥5 with usual FSS item-response mean scoring). Although a ≥4 cut-off was originally proposed in FSS development (Krupp et al., 1989), a ≥5 cut-off is usually chosen in more recent studies due to the ≥4 cut-off overestimating fatigue cases in the general population (Lerdal, Wahl, Rustoen, Hanestad, & Moum, 2005). Hadjimichael et al. (2008) found the "severe fatigue" group was somewhat older, d = 0.25, 95% CI [0.20, 0.29]; more likely to be male (29.4% versus 24.2%); less likely to be employed (29.0% versus 54.6%); educated to below US Associate degree level (61.5% versus 52.0%); and older at MS onset, d = 0.16, 95% CI [0.11, 0.20]. All these findings were highly statistically significant (p < .0001); however most effect sizes were small.

In a survey study of pwMS (n = 368), age was associated with FSS score (b = 0.16) in a multivariate model including gender, education, years since onset, and disease course (Lerdal et al., 2003); however, when the sample was split into those with RRMS/SPMS or PPMS, age was only statistically significant in the PPMS group (b = 0.47) and not the RRMS/SPMS group (b = 0.07). Lerdal et al. (2003) found no association between gender and FSS score. Hadjimichael et al. (2008) did not perform subset analysis by MS-type, but PPMS diagnosis was more prevalent in the "mild/moderate fatigue" group (14.0%) than the "severe fatigue" group (7.6%).

Other studies have failed to find any association between age and fatigue. Chwistiak et al. (2005) split a large sample into those with (n = 177) and without (n = 542) "disabling fatigue" based on a ≥15 cut-off on the MFIS. The study found no group differences for age, d = -0.03, 95% CI [-0.20, 0.14]; prevalence of males (23.7% versus 21.6%); or ethnicity (91.0% Caucasian versus 91.9%). However, education and employment levels were both significantly lower in the "disabling fatigue" group. Flachenecker et al. (2002) found no association ("r = 0") between age and fatigue (FSS) in 151 pwMS.

In summary, associations between age and fatigue have been inconsistently reported and, where statistically significant, only small effect sizes were evident. Although higher proportions of males had severe fatigue in one study (Hadjimichael et al., 2008), the difference was only small and other studies have failed to replicate the finding.

1.2.3 Primary and secondary sources of fatigue

There are many important sources of fatigue in MS, and the pathology of MS fatigue is complex and poorly understood. The MS Council for Clinical Practice Guidelines (1998) suggested fatigue in MS emanates from both primary (centrally-mediated disease factors) and secondary (all other) sources of fatigue: a model of fatigue that has since been employed in a review examining the origins of MS fatigue (Kos et al., 2008). Strober and Arnett (2005) demonstrated sleep disturbance (r = .49) accounted for 24% of variance in MS fatigue (n = 77), with a further 10% attributable to depressive symptoms (r = .47) and 9% attributable to neurological symptoms (r = .34). Predictors were entered into step-wise regression models in order of their zero-order correlations (r values above) with all three factors statistically significant with independent effects in the final model. Sleep disturbance and depressive symptoms were considered to be secondary consequences of MS, whereas neurological symptoms were direct effects of disease-mediated CNS damage; therefore, Strober and Arnett (2005) demonstrated both primary and secondary factors directly impacting on fatigue experience.

Factors designated as secondary sources of fatigue can be described as secondary consequences of MS or other non-MS-related variables. Torres-Harding and Jason (2005) described several sources of non-specific fatigue, which can be applied here in the context of MS: (1) psychosocial stress; (2) unhealthy lifestyles; (3) physiological states (e.g., sleep deprivation and excessive physical exertion); (4) physical comorbidities; (5) psychiatric comorbidities; and (5) medications. With additional primary sources of fatigue, MS fatigue can be described as a multifaceted symptom experience with multiple contributing factors.

1.2.3.1 Secondary sources of fatigue

1.2.3.1.1 Psychosocial stress

While studies have found the majority of pwMS retain the belief that stress affects their fatigue (Mills & Young, 2008; Mollaoğlu & Üstün, 2009; Stuifbergen & Rogers, 1997), studies are hindered by not defining stress or considering which aspects of stress may be relevant. Furthermore, no study has assessed whether any stress-fatigue association is magnified in MS. Lay use of the word "stress" is inconsistent and can mean any one, or all, of three elements: (1) stressor stimuli; (2) intervening processes and pathways; and (3) stress response (S. Levine, 2005). Mills and Young (2008) found 81.8% of 635 pwMS responding "agree"/"strongly agree" to "stress/anxiety worsens fatigue" and, similarly, Mollaoğlu and Üstün (2009) reported 90.0% of 120 pwMS responded "Yes" to the question "Does experiencing stressful situations increase your fatigue?" Neither study defined the stress concept for participants, so it is unclear which facets of stress are thought to exacerbate fatigue. Stuifbergen and Rogers (1997) also found "psychological stress related to family, work, socioeconomic, and other emotionally laden problems" (p.6) was a commonly-described "antecedent" factor of fatigue in semi-structured interviews with 13 pwMS. However, very similar qualitative information about the stress-fatigue relationship has been described elsewhere in healthy individuals (Gledhill, 2005), with "stress, induced by lifestyles, professional overwork, family worries, socio-relational difficulties or transport difficulties" (p.304) believed to induce fatigue.

The transactional model of stress and coping (Lazarus & Folkman, 1984) conceptualises stress as a dynamic bidirectional transaction between person and environment, mediated by cognitive appraisal and coping. Therefore, all environmental stimuli have the potential to provoke a stress response (physiological, behavioural, and emotional), but the magnitude and type of response is dependent upon how stimuli are processed cognitively (Lazarus & Folkman, 1984). The association between perceived stress and MS fatigue has only been examined in one study (Trojan et al., 2007). This cross-sectional investigation of 53 pwMS (37 RRMS, 16 SPMS) measured ratings on the Perceived Stress Scale (PSS; S. Cohen, Kamarck, & Mermelstein, 1983), the FSS, and the physical and mental fatigue subscales of the Multidimensional Fatigue Inventory (MFI; Smets, Garssen, Bonke, & De Haes, 1995). While PSS scores

were correlated with FSS scores, r = .34, 95% CI [.07, .56], the association was no longer statistically significant in an age-adjusted multivariate model including neurological disability. Stress was associated with mental fatigue, r = .43, 95% CI [.17, .62], but not physical fatigue, r = .21, 95% CI [-.06, .46]. While not a predictor of physical fatigue in multivariate modelling, stress was the only statistically significant predictor of mental fatigue, β = 0.25, 95% CI [0.01, 0.49], in an age-adjusted multivariate model (R^2 = .24) including all measured variables (neurological disability, pain, MS self-efficacy, sleep quality).

The limited evidence regarding the stress-fatigue relationship in MS prevents causal inferences; fatigued individuals may appraise situations as more stressful due to fatigue restricting coping resources. It is also unclear whether stress impacts fatigue in MS differently to other populations. In a large general population sample (n = 3438), perceived stress was associated with fatigue (β = 0.30) in a model (R^2 = .28) adjusting for health status, age, gender, and relationship status (Kocalevent, Hinz, Brahler, & Klapp, 2011). Similar associations (β = 0.38) were also found elsewhere in a smaller (n = 325) general population (Michielsen, Willemsen, Croon, de Vries, & Van Heck, 2004). MS fatigue may be sensitive to psychosocial stress, but there is currently no evidence to confirm the causal nature of the relationship, to identify whether certain types of stressors are most relevant, or to implicate physiological, behavioural, or emotional stress responses.

1.2.3.1.2 Physical comorbidity

Additional physical morbidities, particularly those with high prevalence of fatigue such as rheumatic disease (Wolfe, Hawley, & Wilson, 1996), can be considered secondary sources of fatigue in MS. As already described, physical comorbidities are prevalent in MS (36.7-77.1%). Typically, pwMS have regular contact with health professionals, so comorbidities are likely to be identified. A condition highly relevant to fatigue is iron-deficiency anaemia (IDA), where fatigue is the principal symptom. Studies have typically reported IDA prevalence at 2.5-3.5% in MS (Horton, Rudick, Hara-Cleaver, & Marrie, 2010; Kang, Chen, & Lin, 2010) compared to a 0.6% rate in the general population (Kang et al., 2010). PwMS are regularly screened for IDA as part of routine blood tests examining haemoglobin, ferritin and iron counts, and screening for thyroid disorders (MS nurse, personal communication). Haemoglobin and iron

serum levels are no different between pwMS and healthy individuals without an IDA diagnosis (Sfagos et al., 2005).

1.2.3.1.3 Psychiatric comorbidity

Lifetime prevalence of MDD in MS is around 50% (Goldman Consensus Group, 2005) and associations between fatigue and depressive symptoms in MS are frequently high; possibly due to fatigue being a common symptom of depression (Bol et al., 2009). However, moderate-high associations have been reported even where correcting for somatic overlap in depression scales ($r = .58$, Kroencke et al., 2000). Fatigue and depression can also have similar characteristics, including lacking motivation and sleep disturbance (Krupp et al., 1988).

Many studies have demonstrated positive associations between depressive symptoms and fatigue (Bakshi et al., 2000; Chwastiak et al., 2005; Flachenecker et al., 2002; Kroencke et al., 2000), but in smaller studies there has been no significant relationship (Krupp et al., 1988; Krupp et al., 1989). Longitudinal studies have not elucidated the causal direction of the relationship, with positive associations between depressive symptoms and change in fatigue over time (R. F. Brown et al., 2009; Schreurs, de Ridder, & Bensing, 2002; Téllez et al., 2006), but also fatigue predicted future depressive symptoms (R. F. Brown et al., 2009).

1.2.3.1.4 Medication

A study of 320 pwMS found no difference in fatigue levels between those receiving and not receiving immunomodulating drugs including interferon-beta 1α, interferon-beta 1β, or glatiramer acetate, $OR = 0.95$, $p = .85$, 95% CI [0.6, 1.6] (Putzki, Katsarava, Vago, Diener, & Limmroth, 2008). There was also no statistically significant risk associated with immunosuppressant drugs like mitoxantrone and azathioprine, $OR = 1.34$, $p = .38$, 95% CI [0.7, 2.6], although these DMTs are not licensed for MS treatment in the UK.

Modafinil and amantadine are two drugs infrequently prescribed "off-label" (unlicensed) in the UK to pwMS with severe fatigue (NICE, 2013). Modafinil is most frequently used in narcolepsy, and is thought to work on brain regions which regulate wakefulness such as the hypothalamus; however, its efficacy has been shown to be no better than placebo for MS fatigue

(Stankoff et al., 2005). Amantadine is an antiviral drug used in Parkinson's disease, but similarly was no better than placebo for MS fatigue in an RCT (Krupp et al., 1995). However, these drugs do seem to work anecdotally for some individuals (Bakshi, 2003).

1.2.3.1.5 Sleep deprivation or disturbance

Trojan et al. (2007) demonstrated that Pittsburgh Sleep Quality Index (Buysse, Reynolds III, Monk, Berman, & Kupfer, 1989) score predicted general fatigue on the MFI, β = 0.35, 95% CI [-0.01, 0.71], in a multivariate model (R^2 = .39) accounting for neurological disability, age, pain, perceived stress, and physical activity. A self-report survey including the Medical Outcomes Study Sleep Scale (Hays & Stewart, 1992) found 51.5% of 1063 pwMS had moderate-severe sleep disturbance (Bamer, Johnson, Amtmann, & Kraft, 2008). While this study had a very low response rate (19%), it corroborated a previous report of sleep-disturbance in 54% of 28 pwMS (Tachibana et al., 1994). However, in a study employing validated mobile polysomnography, 96% of fatigued pwMS (n = 26; MFIS ≥45) and 60% of non-fatigued pwMS (n = 40) met criteria for a clinically relevant sleep disorder (Veauthier et al., 2011). The rates observed by Veauthier et al. (2011) seem very high, and may be biased by individuals with disturbed sleep probably being more willing to take part.

Restless leg syndrome is a sleep-related movement disorder involving leg sensations that begin or worsen during rest (Earley, 2003). A large multi-centre survey of 861 pwMS (649 RRMS, 50 PPMS, 162 SPMS) and 649 healthy individuals found prevalence of restless leg syndrome (using standard diagnostic criteria) was far more prevalent in pwMS (19.0%) than controls (4.2%), OR = 5.4, 95% CI [3.56, 8.26] (Manconi et al., 2008). In addition, 24.5% of the MS group reported experiencing insomnia in the prior 6 months. A review of sleep disturbance and fatigue in MS described several MS-related factors contributing to sleep quality, including bladder dysfunction, pain, leg spasms, and depression (Kaminska, Kimoff, Schwartzman, & Trojan, 2011). MS medications may also be contributing factors (Kaminska et al., 2011), but this has not been empirically examined. Pain has been shown to predict future fatigue in MS (Patrick et al., 2009), although whether this association is via sleep disturbance is an untested mechanism.

1.2.3.1.6 Cognitive and behavioural factors

Cognitive interpretations and behavioural responses to symptoms can influence MS fatigue experience (Skerrett & Moss-Morris, 2006). These factors include the tendency to attribute all somatic stimuli to MS (β = 0.33), all-or-nothing (push themselves then crash) behaviours (β = 0.20), and avoidance behaviours (β = 0.25) which were all associated with mental fatigue in cross-sectional multivariate regression analysis controlling for remission status and neurological disability (Skerrett & Moss-Morris, 2006). All-or-nothing behaviours were also associated with physical fatigue (β = 0.19) controlling for the same disease-related variables. Van Kessel and Moss-Morris (2006) presented a cognitive-behavioural model of MS fatigue proposing central disease mechanisms, such as inflammation and demyelination, trigger fatigue which is then maintained as a chronic symptom (a "vicious cycle of fatigue") by maladaptive cognitions and behaviours. An 8-week cognitive behavioural therapy (CBT) intervention for MS fatigue based on this model of fatigue was compared to relaxation therapy in an RCT (van Kessel et al., 2008). The CBT group (n = 35) reported greater reductions in fatigue severity than the relaxation therapy group (n = 37), although clinically significant effects were apparent for both interventions (CBT: d = 3.03; relaxation therapy: d = 1.83).

1.2.3.2 Primary sources of fatigue

Primary fatigue is fatigue caused by centrally-mediated disease processes. Several hypotheses have been proposed, relating to the immune, nervous, and endocrine systems.

1.2.3.2.1 Immune factors

Several studies have demonstrated alterations in pro-inflammatory cytokine profiles in pwMS who do and do not experience fatigue. Flachenecker et al. (2004) divided a group of 37 pwMS (29 RRMS, 8 SPMS) into those with (n = 26) and without (n = 11) severe fatigue based on the ≥4 FSS cut-off. While interferon gamma (IFN-γ) and interleukin-10 (IL-10) levels were no different between groups, tumor necrosis factor alpha (TNF-α) levels were greater in those with fatigue than without. This finding was later corroborated by Heesen et al. (2006) who found elevated TNF-α and IFN-γ serum levels in a group of individuals with fatigue (n = 15; 6 RRMS, 8 SPMS, 1 PPMS) compared to a group without (n = 15; 11 RRMS, 2 SPMS, 2 PPMS), based on the ≥5 FSS

cut-off. Another study measured fatigue in 44 pwRRMS using the Wurzburg Fatigue Inventory for MS (Flachenecker et al., 2006) and, while finding no association with TNF-α producing T-cells, there was an association with IFN-γ producing T-cells (*R* = 0.44) (Gold et al., 2011). However, another study found no association between IFN-γ producing T-cells and fatigue in 38 pwMS (16 RRMS, 9 SPMS, 13 PPMS) (Giovannoni, Thompson, Miller, & Thompson, 2001).

While mixed findings are presented regarding the role of pro-inflammatory cytokines in MS fatigue, the role of MS type may be important; cytokine profiles (Ysrraelit, Gaitán, Lopez, & Correale, 2008) and fatigue levels (Patrick et al., 2009) can be different across MS types. Only the Gold et al. (2011) study above recruited a homogeneous MS sample or accounted for MS type. Strong associations in cross-sectional studies between specific inflammatory markers and MS fatigue may be difficult to observe as inflammatory cytokines fluctuate markedly with, and independent of, disease activity in MS (Induruwa et al., 2012).

1.2.3.2.2 CNS factors

CNS lesion load may be relevant to MS fatigue, but evidence is mixed. Colombo et al. (2000) demonstrated an association between FSS scores and lesion load on MRI (*rho* = .5) in 30 pwRRMS without MDD and with low levels of physical disability. Higher lesion loads were also found in the higher parietal lobe, internal capsule, and periventricular trigone in a subgroup with MS fatigue (*n* = 15). In another study of 222 pwMS (MS type unclear) with low levels of disability, T1 and T2 lesion loads were higher in fatigued (FSS ≥5; *n* = 197) than non-fatigued (FSS ≤4; *n* = 25) groups (Tedeschi et al., 2007). However, in a study of 71 pwMS (50 RRMS, 21 SPMS), FSS score was not associated with regional or global lesion load (Bakshi et al., 1999) and, in a smaller study of 28 pwMS (type unclear), those with fatigue had similar lesion loads than those without fatigue (Codella et al., 2002).These studies suggest heterogeneous MS groups may mask an association between fatigue and lesion load seemingly more apparent in RRMS.

When studies account for confounding factors such as depression and physical disability, lesion load seems relevant to fatigue (Kos et al., 2008); however, they provide little information about causation. A more recent study

examined corpus callosum atrophy and hypothesised that, as a large and important region commonly affected by MS, it may be implicated in fatigue (Yaldizli et al., 2011). Seventy pwRRMS with low levels of depressive symptoms were split into those with (n = 28) and without (n = 42) fatigue based on the FSS ≥4 cut-off, and corpus callosum atrophy was computed from MRIs at diagnosis and time of study. Mean corpus callosum atrophy since diagnosis (M = 4.8 years) was greater in the fatigued group (-21.8%) than non-fatigued (-12.1%); after adjusting for time since diagnosis and disability, atrophy rate was -1.1% per year in the fatigued group and -0.6% per year in the non-fatigued group. A similar study of 134 pwRRMS found an association between fatigue and progressive brain atrophy over an eight-year follow-up period (Marrie, Fisher, Miller, Lee, & Rudick, 2005). A review by Vucic, Burke, and Kiernan (2010) suggested other mechanisms of fatigue involving the CNS may include sodium channel dysfunction, increased cortical activation, and reduced glucose metabolism in cortical motor and basal ganglia regions.

1.2.3.2.3 Endocrine factors

With circulating levels of pro-inflammatory cytokines being elevated in MS compared to healthy individuals (Ysrraelit et al., 2008), the anti-inflammatory properties of cortisol, its role in stress regulation, as well as its involvement in energy metabolism (Sapolsky et al., 2000) makes cortisol a potentially relevant endocrine biomarker in MS fatigue. With cortisol as its "end product", the HPA axis is an important focus within this thesis. The remainder of this chapter will address the HPA axis in MS, culminating in its potential role in MS fatigue.

1.3 Hypothalamic-pituitary-adrenal axis

The HPA axis represents a hormonal cascade pathway incorporating the hypothalamus, pituitary gland, and adrenal glands. The paraventricular nuclei of the hypothalamus synthesise and release the neuropeptides corticotrophin-releasing hormone (CRH) and arginine-vasopressin (AVP). These hormones stimulate the production and secretion of adrenocorticotropic hormone (ACTH) by the anterior pituitary gland. Once secreted, ACTH acts on the adrenal cortex to promote the release of steroid hormones called glucocorticoids (cortisol in humans) into circulation (Sapolsky et al., 2000). The HPA axis is a self-

regulatory system via negative feedback, with cortisol binding at receptors inhibiting synthesis of CRH and ACTH (D. B. Miller & O'Callaghan, 2002).

The HPA axis is a psychophysiological system central to *allostasis*: a term describing the active dynamic processes by which organisms achieve physiological stability through change rather than constancy (Sterling & Eyer, 1988). However, physiological arousal may come at a cost (*allostatic load*) if sustained such as under chronic stress. Allostatic load provides a theoretical perspective linking stress to health consequences via wear and tear caused by inadequate, inappropriate, or sustained physiological responses (McEwen, 1998, 2004). As well as vital to supporting normal physiological functioning by regulating immune processes and energy metabolism (Dickerson & Kemeny, 2004; Sapolsky et al., 2000), the HPA axis is also highly sensitive to acute psychological stress (Dickerson & Kemeny, 2004; G. E. Miller, Chen, & Zhou, 2007). Dickerson and Kemeny (2004) presented a meta-analysis investigating cortisol responses to acute laboratory-induced psychological stressors in healthy adults. Eligible studies included stressors lasting less than one hour and excluded physical stressors such as physiological challenge tests. Typical acute stressors included public speaking tasks or emotion-induction (for example, a provocative film). The review found 208 studies and a small-medium average effect size, d = 0.31, 95% CI [0.22, 0.40], indicating an average 0.31 *SD* rise in cortisol from baseline following stressors. These findings have translated into everyday life, with statistically significant elevations in cortisol following momentary social stress (items included "I would rather be alone") and recent "unpleasant" events in a healthy female sample (Jacobs et al., 2007).

1.3.1 Cortisol

The "end product" of human HPA axis activity is cortisol, which circulates in one of two forms: protein-bound cortisol and free cortisol. Free cortisol is acknowledged as the physiologically-active form available to bind at target location receptors and activate processes (Williams & Dluhy, 2006). The acute effects of cortisol are generally adaptive and cortisol has a principal function in regulating immune activity. Typically, cortisol acts to suppress pro-inflammatory cytokines, preventing overshoot of inflammatory immune responses to, for example, infectious stressors (Sternberg, 2006). This anti-

inflammatory characteristic is demonstrated by high-dose synthetic corticosteroid treatments alleviating symptom exacerbations in inflammatory diseases such as MS and rheumatoid arthritis (Gorter et al., 2010; Milligan et al., 1987).

Cortisol mobilises energy reserves by promoting the breakdown of fats and proteins into glucose (Sapolsky et al., 2000). While catecholamines (adrenaline and noradrenaline) provide instant energy "boosts", cortisol increases the availability of usable carbohydrates, providing an opportunity to restore energy levels in the longer-term. The effects of catecholamines on blood pressure and heart rate are further potentiated by cortisol (Sapolsky et al., 2000). Cortisol is also thought to increase appetite (Dallman et al., 2004) and support memory formulation for situations involving emotional arousal, such as threats (Abercrombie, Speck, & Monticelli, 2006).

Under normal conditions, cortisol secretory activity (CSA) follows a 24h circadian rhythm characterised by slowly increasing levels during the second half of sleeping with a strong rise upon awakening (38-75% increase, peaking within 20-30 minutes), followed by a steady decline over the rest of the day with lowest levels during the first part of the night (Fries et al., 2009). The waking rise in cortisol is termed the *cortisol awakening response* (CAR) or other synonymous phrase (J. C. Pruessner et al., 1997) and the steady daytime decline is the *diurnal cortisol slope* (DCS). Cortisol is secreted throughout the day in a pulsatile fashion with periods of HPA axis inhibition and activation (E. A. Young, Abelson, & Lightman, 2004).

A large number of psychosocial and methodological factors have been associated with the CAR (Chida & Steptoe, 2009) including variables such as time of awakening, sleep quality, and demands of the upcoming day (Edwards, Evans, Hucklebridge, & Clow, 2001; Schlotz et al., 2004; Wüst et al., 2000). The CAR is an important marker of basal (unstimulated) HPA axis activity and was observed in 100% of 16 healthy participants in laboratory conditions (Wilhelm, Born, Kudielka, Schlotz, & Wüst, 2007) but typically around 75% in everyday life (Wüst et al., 2000). This difference may be due to poorer compliance in everyday life, but the novelty of the laboratory environment may also invoke additional arousal upon awakening (Wilhelm et al., 2007). Less research has been carried out on the DCS, but a flat DCS has been associated with factors such as greater age and depression (Heaney, Phillips, & Carroll,

2010; Sjögren, Leanderson, & Kristenson, 2006), and is often an artefact of elevated evening cortisol (for example, Deuschle et al., 1997).

The allostatic load model presents cortisol as a primary stress mediator of physiological dysfunction leading to disease or disorder (McEwen & Seeman, 1999). Free cortisol acts on target cells via glucocorticoid receptors and, over time, glucocorticoid receptors may become desensitised by prolonged exposure, diminishing the regulatory control of cortisol over immune inflammation and HPA axis self-regulation (G. E. Miller, Cohen, & Ritchey, 2002). HPA axis dysfunction can manifest as hyperactive output (hypercortisolism) or attenuated output and reactivity (hypocortisolism). Hyperactivity of the HPA axis has been associated with MDD (Pariante & Lightman, 2008), type II diabetes (Rosmond, 2003) and cardiovascular disease (Whitworth, Williamson, Mangos, & Kelly, 2005), while hypocortisolism has been associated with conditions such as fibromyalgia and CFS (Fries, Hesse, Hellhammer, & Hellhammer, 2005; Heim, Ehlert, & Hellhammer, 2000).

1.3.1.1 Cortisol measurement

Current techniques for measuring circulating cortisol levels include sampling of saliva, serum (blood), urine, and hair (Gow, Thomson, Rieder, Van Uum, & Koren, 2010; Kirschbaum & Hellhammer, 1989; A. Levine, Zagoory-Sharon, Feldman, Lewis, & Weller, 2007). Saliva sampling enables non-invasive repeated measurement of free cortisol, and has excellent affinity with serum cortisol (r = .85) (Gallagher, Leitch, Massey, McAllister-Williams, & Young, 2006) while not requiring venepuncture (piercing a vein to draw blood). Urinary cortisol provides only a summary measure of cortisol output over 24h, and hair sampling is a newer technique reserved for longer-term retrospective quantification of cortisol levels (Russell, Koren, Rieder, & Van Uum, 2012). Saliva assessments have enabled researchers to design cortisol studies in everyday life which are more generalizable to daily experience (Schlotz, 2011). Descriptions and rationale for specific composite markers of CSA will be presented in Chapter 2.

1.3.2 HPA axis activity in MS

Given MS is an inflammatory disease potentially sensitive to stress, and relapses are alleviated by exogenous corticosteroids (Tremlett, Luscombe, &

Wiles, 1998), the relevance of HPA axis activity to MS is a potentially important focus for MS research. Pharmacological challenge tests such as the dexamethasone suppression test (DST) (Liddle, 1960) or combined dexamethasone/CRH (Dex/CRH) (Heuser, Yassouridis, & Holsboer, 1994) test have been used to examine HPA axis function in MS, but results have been mixed. More recently, some studies have examined basal (unstimulated) CSA in MS to examine markers such as the CAR and DCS. These studies are now reviewed in the sections below; for details of the standard procedure and rationale for the DST and Dex/CRH test, see Appendix B.

1.3.2.1 Pharmacological challenge tests

Dexamethasone acts by mimicking the actions of cortisol to suppress further ACTH and cortisol production. A less sensitive test, the DST was generally only performed in earlier studies in MS and results have been mixed. While some studies have reported low rates of non-suppression in the DST (0-12%) (Then-Bergh, Kümpfel, Trenkwalder, Rupprecht, & Holsboer, 1999; Wei & Lightman, 1997) others have reported dysfunction (33-62%) (Heesen, Gold, Raji, Wiedemann, & Schulz, 2002; Reder, Lowy, Meltzer, & Antel, 1987; Ysrraelit et al., 2008). Of note, studies indicating low non-suppression rates deviated from the standard DST protocol, delaying morning samples by 30-60 minutes. Healthy control groups, where present, have demonstrated non-suppression rates between 0% (Then-Bergh et al., 1999; Ysrraelit et al., 2008) and 19% (Reder et al., 1987). MS type may again be relevant; pooling data across studies which distinguished between MS types revealed somewhat lower non-suppression rates in RRMS (33.8% of 74) and SPMS (38.7% of 62) than PPMS (50.0% of 46).

The determination of cortisol, a pulsatile hormone, by a single sample is a major criticism of the DST and the Dex/CRH test is now typically preferred (Heuser et al., 1994). Then-Bergh et al. (1999) reported very large cortisol hyper-response ($\Delta max_{cortisol}$) in PPMS (n = 5) compared to controls (n = 29), d = 2.37, 95% CI [1.28, 3.34], but both RRMS (n = 30), d = 0.19, 95% CI [-0.30, 0.67], and SPMS (n = 14), d = 0.38, 95% CI [-0.24, 0.99], were not significantly different to controls. While evidence described so far would suggest greater HPA axis hyperactivity in PPMS, another study found no difference between a PPMS group (n = 13) and controls (n = 11), d = -0.05, 95% CI [-0.85, 0.75],

(Heesen, Gold, et al., 2002). Small sample sizes prohibit drawing any firm conclusions, particularly in PPMS.

Of most relevance to this thesis is RRMS; but again, studies with RRMS groups or subgroups reported mixed results (Fassbender et al., 1998; Grasser, Möller, Backmund, Yassouridis, & Holsboer, 1996; Heesen, Gold, et al., 2002; Then-Bergh et al., 1999). Cortisol hyper-responses were observed in RRMS groups compared to healthy control groups by Grasser et al. (1996) (n = 19), d = 0.81, 95% CI [0.13, 1.46], and Fassbender et al. (1998) (n = 23), d = 0.93, 95% CI [0.25, 1.57], but no substantial differences were observed by either Then-Bergh et al. (1999) (n = 30), d = 0.19, 95% CI [-0.30, 0.67], or Heesen, Gold, et al. (2002) (n = 8), d = 0.16, 95% CI [-0.76, 1.06].

Level of disease activity may be responsible for different results. Schumann et al. (2002) found attenuated cortisol responses to the Dex/CRH test in pwMS with Gd+ lesions on MRI compared to those without Gd+ lesions, d = -0.86, 95% CI [-1.47, -0.22], in a mixed-type MS group (35 RRMS, 13 SPMS, 5 PPMS). Number of Gd+ lesions was also associated with cortisol response (r = -.44) and, in RRMS specifically, cortisol hypo-responders had more Gd+ lesions (M = 4.0 ±5.0) than hyper-responders (M = 1.1 ±1.8). The inference was that HPA axis hyper-response is a protective mechanism, probably via inflammation restraint. However, the study does not allow causal inferences, and Fassbender et al. (1998) found the opposite: cortisol hyper-response to the Dex/CRH test in those with Gd+ lesions, d = 0.95, 95% CI [0.02, 1.81], with a positive association between white blood cell count in cerebrospinal fluid (an indicator of CNS inflammation) and cortisol hyper-response (r = .63). Fassbender et al. (1998)'s results suggest HPA axis dysfunction via chronic exposure to inflammatory mediators affecting glucocorticoid receptor sensitivity. Both hypotheses require further examination in longitudinal studies. Fassbender et al. (1998) reported 4 of the 24 in the MS sample met DSM-III criteria for MDD. Hyper-response to the Dex/CRH test is considered a valid biomarker of MDD (Ising et al., 2007), potentially confounding any attribution of hyper-response to MS disease activity. Fassbender et al. (1998) reported a statistically significant association between cortisol hyper-response and depressive symptoms (r = .56), possibly driven by individuals with MDD given other studies have found smaller associations (r < .20) (Heesen, Gold, et al., 2002; Then-Bergh et al., 1999).

Schumann et al. (2002) did not report psychiatric morbidity as an exclusion criterion in their study, and depressive symptoms were not reported.

1.3.2.2 Cortisol reactivity to stress in MS

Acute psychosocial stressors typically provoke increased cortisol levels from baseline, particularly when the stressor is a public speaking/cognitive task combination (Dickerson & Kemeny, 2004). Two studies included in the Dickerson and Kemeny systematic review were conducted in MS (Ackerman, Martino, Heyman, & Moyna, 1996; Heesen, Schulz, et al., 2002) and, although stressor presentations were different, both included motivated performance tasks with socio-evaluative threat (public speaking). Both studies suggested HPA axis hypo-reactivity to acute psychosocial stress.

Heesen, Schulz, et al. (2002) implemented a 45min stressor incorporating a mental arithmetic task, Stroop task, and video-recorded public-speaking task. An RRMS group was randomly assigned to a "stress" group (n = 20) or "no-stress" group (n = 15), while a healthy control group (n = 15) also received the stressor. Exact stressor details were not reported, but the stressor was substantiated by elevated heart rate immediately after the combined stressor. The "no-stress" group protocol was not reported. Cortisol levels were measured in blood before (t0), immediately after (t1), and one hour (t2) following stressor completion. All stressor protocols began at 0830h. A statistically significant effect showed cortisol reduced over time, with no statistically different group or group*time effects. The MS-stress group showed a non-significant reduction in cortisol from t0 to t1, d = -0.57, 95% CI [-1.27, 0.15], but cortisol continued to reduce to t2, d = -1.21, 95% CI [-1.93, -0.43]. However, reductions in cortisol from t0 to t2 in the MS "non-stress" group, d = -0.49, 95% CI [-1.31, 0.38], and the healthy control group, d = -0.94, 95% CI [-1.79, -0.03], suggest a systematic problem in the methodology. At 0830h, cortisol levels are particularly dependent on time of awakening (not reported), and it is possible that many CARs had peaked and cortisol levels were reducing following a secretion pulse during the study protocol.

Ackerman et al. (1996) presented similar findings, but again performed the study protocol in the early morning (0700h – 1000h) without considering time of awakening. The study included 25 pwMS (RRMS or SPMS; ratio unclear)

and 25 healthy controls, and the stressor involved a short (5min) video-recorded public speaking task. After insertion of a catheter in the forearm (and 30min rest), the task was performed. Blood samples were obtained at baseline (t0), 5min (t1, end of stress), 20min (t2) and 60min (t3) following stressor onset. Results showed a small, statistically significant, increase in cortisol from t1 (M = 16.4) to t2 (M = 17.9), but no significant difference between onset and any later measurement (t0 cortisol > t1 cortisol). Of note, the study found the stressor did not increase heart rate or blood pressure.

Both Heesen, Schulz, et al. (2002) and Ackerman et al. (1996) suggest minimal or reduced cortisol levels in response to stress, implying hypo-reactivity. However, this conclusion should be made with extreme caution due to the methodological problems described. Observing reactivity to everyday life stressors (daily hassles) may provide the best opportunity for further investigate as there are now ethical considerations in inducing acute stressors on a clinical population where stress is associated with negative health outcomes (Artemiadis et al., 2011; Mohr et al., 2004).

1.3.2.3 Basal cortisol secretory activity in MS

Basal (unstimulated) CSA has been examined in MS using both urinary cortisol and salivary cortisol measurement. Michelson et al. (1994) reported 24h urinary cortisol levels higher in 12 pwMS (10 RRMS, 3 progressive MS) than 12 healthy individuals, d = 1.05, 95% CI [0.16, 1.86]. A more comprehensive study reported greater 24h urinary cortisol in all MS-type subgroups compared to respective age- and sex-matched healthy control groups: remission-phase RRMS (n = 34), d = 0.68, 95% CI [0.23, 1.13]; relapse-phase RRMS (n = 58), d = 2.55, 95% CI [1.86, 3.18]; PPMS (n = 40), d = 2.03, 95% CI [1.42, 2.58]; and SPMS (n = 41), d = 1.92, 95% CI [1.33, 2.46]. These results demonstrate cortisol is elevated across all MS types, but is lowest in the remission phase of RRMS where inflammation and neurodegeneration levels are lower than in remission-phase RRMS or progressive MS types, respectively.

In more recent years, three case-control studies have explored CSA in everyday life in MS using salivary cortisol assessments (Gold et al., 2010; Kern et al., 2013; Kern et al., 2011) and, with the exception of Kern et al. (2013), recruited homogeneous remission-phase RRMS groups. Gold et al. (2010) implemented a relatively infrequent sampling design, with samples taken at

awakening, 1600h, and 2100h over 2 consecutive days, in an RRMS group (n = 29) and a healthy control group (n = 20). After aggregating across days, there were no statistically significant differences between groups at awakening, 1600h, or 2100h (data not provided), although flatter DCS in the RRMS group was close to statistical significance (p = .08). After splitting the RRMS group into those with (n = 8) and without (n = 21) mild depression based on the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996) ≥14 cut-off, those in the BDI-high group had higher 2100h cortisol levels than healthy controls and a flatter DCS. However, this study did not exclude comorbidities, and seven individuals in the RRMS group (two in the BDI-high, five in the BDI-low) were taking antidepressants. Antidepressants have been shown to normalise Dex/CRH hyper-responses in MDD (Ising et al., 2007). The study did not obtain or exclude clinical diagnoses of MDD, so it is possibility that MDD was present within the low-BDI group, particularly given antidepressant-use was apparent. The BDI was shown to yield a false-positive rate of 30% for MDD in a previous study in MS (Sullivan, Weinshenker, Mikail, & Bishop, 1995).

Gold et al. (2011) later collected salivary cortisol assessments in two female-only remission-phase RRMS groups: with (n = 10) or without MDD (n = 34). MDD was determined by a clinical psychologist (DMS-IV criteria), and individuals with a medical history of endocrine abnormalities (e.g., diabetes mellitus) or other psychiatric diagnoses were excluded. Cortisol was sampled over 2 consecutive days at awakening, 1100h, 1500h, 2000h, and 2300h to measure daily cortisol output, and at awakening (T0), 15 minutes after awakening (T15), T30, T45, and T60 to measure the CAR. Results showed the RRMS+MDD group had higher daily cortisol output than the RRMS group, d = 0.74, 95% CI [0.00, 1.45], along with a flatter DCS. No difference was found for the CAR between groups.

Both studies by the Gold research group implied HPA axis hyperactivity in MS, if present, can be attributed to the high prevalence of MDD in MS rather than to the disease itself. With HPA axis hyperactivity frequently observed in MDD (Knorr, Vinberg, Kessing, & Wetterslev, 2010; Pariante & Lightman, 2008) and a 50% lifetime prevalence of MDD in MS (Goldman Consensus Group, 2005), this appears a plausible hypothesis. However, the relatively small sample size in the Gold et al. (2010) study suggests flatter DCS in the RRMS group (p = .08) may (or may not) have been statistically significant with more

participants. Ysrraelit et al. (2008) stated that depressed individuals did not differ from non-depressed in 24h urinary cortisol or DST outcomes; although it was unclear how MDD was determined. Gold et al. (2005) found pwMS who were hyper-responders to the Dex/CRH test experienced greater disability progression over the next 3 years, suggesting disease mechanisms is also relevant to HPA axis activity.

Kern et al. (2011) employed a median split on the BDI-II (median = 7) to examine the role of depression in CSA measured by saliva in 32 pwRRMS and 16 age- and sex-matched healthy individuals. Saliva samples were provided over 2 days (within 1 week) at T0, T30, T45, and T60, 1500h and 2200h. Similar to Gold et al. (2010), the study reported differences in CSA compared to controls in the BDI-high group, but in the CAR rather than daily output. While the RRMS group had elevated CAR output, d = 0.76, 95% CI [0.13, 1.36], compared to controls, this difference was only statistically significant for the BDI-high group, d = 1.01, 95% CI [0.25, 1.72], and not the BDI-low group, d = 0.55, 95% CI [-0.17, 1.24]. However, the inference that depression is driving this difference is questionable. The median of 7 on the BDI-II is well within the "minimal depression" range (scores 0–13) for the scale (Beck et al., 1996) and has little clinical relevance to MDD. Furthermore, statistical power was a clear problem after the median split, with large 95% CI range potentially masking a moderate-sized average effect even in the BDI-low group. It was also not clear whether any participants in this study were taking antidepressants.

The most recent study recruited 55 pwRRMS (in remission), 22 pwSPMS, and 34 healthy individuals (Kern et al., 2013), and used an identical saliva sampling protocol to the Kern et al. (2011) study. Depressive symptoms were measured by the Centre for Epidemiological Studies Depression (CES-D) Scale (Radloff, 1977). The study reported elevated CAR output in RRMS compared to healthy controls, d = 0.57, 95% CI [0.13, 1.00], and CES-D score was not associated with CAR (r = .02). A median split by CES-D was not performed, and the authors described levels of depression in the RRMS group as "mild". The SPMS group did not differ from controls for any cortisol measure. However, it seemed from the report that comorbidity was not an exclusion criterion.

In summary, studies examining basal CSA in MS have described heightened CSA in relapse-phase RRMS, SPMS, and PPMS (Ysrraelit et al., 2008)

and remission-phase RRMS (Gold et al., 2010; Kern et al., 2013; Kern et al., 2011; Ysrraelit et al., 2008); however some studies have suggested heightened CSA in MS may be driven by high levels of depression within the population (Gold et al., 2010; Gold et al., 2011; Kern et al., 2011). Very recently, a salivary cortisol study has demonstrated heightened CSA in a sample with low levels of depressive symptoms (Kern et al., 2013) and a previous 24h urinary cortisol study described no difference in cortisol output between depressed and non-depressed pwMS (Ysrraelit et al., 2008). Salivary cortisol studies enabled the assessment of CSA in everyday life, with clear ecological validity. However, findings are mixed and several factors contribute. None of the case-control studies by the Gold or Kern groups have explicitly stated the exclusion of important comorbidities whose disease mechanisms or treatments may affect cortisol levels. The attribution of cortisol hyper-secretion in MS to MDD has also been made without clinical diagnoses. Methodologically, no study made any attempt to maximise, measure, or control participant compliance, which was an important omission given only moderate (62%) compliance rates with salivary cortisol protocols reported elsewhere (Broderick, Arnold, Kudielka, & Kirschbaum, 2004).

1.3.3 HPA axis and MS fatigue

As cortisol is also implicated in the mobilisation of energy resources (Sapolsky et al., 2000) it is a candidate for research into MS fatigue. The relationship between fatigue and HPA axis activity has been examined most extensively in chronic fatigue syndrome (CFS), with systematic reviews indicating an association between reduced HPA axis activity and CFS (Cleare, 2003; Papadopoulos & Cleare, 2012); although no study has identified HPA axis hypo-activity as causal in the onset of CFS. There is limited evidence elsewhere examining whether similar associations are present with MS fatigue or fatigue in other clinical populations.

Only three studies have examined the relationship between HPA axis activity and MS fatigue (Gold et al., 2011; Gottschalk et al., 2005; Heesen et al., 2006). Two of these studies examined HPA axis function with the Dex/CRH test (Gottschalk et al., 2005; Heesen et al., 2006) with both splitting MS groups into fatigued and non-fatigued groups based on FSS cut-offs; Gottschalk et al. (2005) employed a lower cut-off (≥4) than Heesen et al. (2006) (≥5).

Heesen et al. (2006) reported no statistically significant differences for either ACTH or cortisol responses in the Dex/CRH test between MS fatigued (n = 15) and non-fatigued (n = 15) groups. However, baseline cortisol (1500h; before CRH stimulation) showed a correlation trend with FSS score (r = .38, p < .08), meaning individuals with less suppression by dexamethasone had higher fatigue scores. The fatigued group also demonstrated significantly greater inflammatory cytokine profiles for both IFN-γ and TNF-α. Similarly, Gottschalk et al. (2005) found no difference between fatigued (n = 15) and non-fatigued (n = 16) MS groups for cortisol responses in the Dex/CRH test, but ACTH responses were greater in the fatigued group. The graphed data suggested a difference in the cortisol response as well, which perhaps would have been statistically significant were in not for the small n. Gottschalk et al. (2005) recruited only pwRRMS, whereas Heesen et al. (2006) had heterogeneous MS groups: fatigued (6 RRMS, 8 SPMS, 1 PPMS) and non-fatigued (11 RRMS, 2 SPMS, PPMS). The role of the HPA axis in MS fatigue may therefore be most prominent in RRMS, but this interpretation is cautious given the different cut-offs categorising fatigue.

The final study looked at the association between fatigue and salivary cortisol in an RRMS sample (Gold et al., 2011), and was described in section 1.3.2.3. Fatigue was measured by the Wurzburg Fatigue Inventory for MS (Flachenecker et al., 2006), but neither the CAR nor the DCS were associated with fatigue. However, a substantial proportion of the sample on which this association was tested also had comorbid MDD; a condition associated with both HPA axis hyperactivity (Pariante & Lightman, 2008) and high levels of fatigue (American Psychiatric Association, 2000).Of other biomarkers measured, only IFN-γ producing T cells (not TNF-α, or Natural Killer Cells) were associated with fatigue (r = .44).

In summary, research examining the relationship between HPA axis and MS fatigue has been limited to small-scale Dex/CRH test studies and one study in everyday life which had several limitations. Pro-inflammatory cytokines were associated with MS fatigue in the studies above, but the number of Gd+ lesions on MRI (the gold standard for observing active inflammation in MS) has had no relationship with MS fatigue previously (Mainero et al., 1999; Marrie et al., 2005). Chronic stress may be involved in both chronic activation of the HPA

axis and in manifesting fatigue, and is an interesting potential mediator of any cortisol-fatigue association.

1.4 Chapter summary and rationale for research project

MS is a chronic inflammatory autoimmune disease of the CNS and has three main types: RRMS, SPMS, and PPMS. MS fatigue appears to differ from fatigue experienced in the general population. It is a complex phenomenon in a complex disease and, accordingly, its definition and measurement lacks consensus. Fatigue is commonly experienced across all MS-types and research investigating primary and secondary sources of fatigue has been presented. Qualitative explorations have suggested fatigue is most severe in the late afternoon and highly sensitive to daily demands in MS (Mills & Young, 2008); however, the diurnal fatigue pattern has not been explicitly examined prospectively in everyday life, and MS fatigue sensitivity to everyday contextual factors has not been explored.

Cortisol is a relevant target for research examining factors contributing to fatigue in MS given its role in the regulation of inflammation, energy metabolism and the stress response system (Sapolsky et al., 2000). Salivary cortisol studies in RRMS have generally suggested HPA axis hyperactivity in RRMS may be attributable to the high levels of depression within the MS population (Gold et al., 2010; Gold et al., 2011; Kern et al., 2011), although the most-recent study (Kern et al., 2013) found larger CARs in an RRMS population with low levels of depressive symptoms. However, all case-control studies examining CSA in everyday life in MS did not report the exclusion of individuals with other health conditions, lacked any means of maximising participant compliance to saliva sampling protocols, and had few sampling days. The relationship between cortisol and fatigue in RRMS is unclear as the only study to previously test this association did so in a sample where 10 individuals (23% of total sample) also had comorbid MDD (Gold et al., 2011).

1.4.1 Research aims

This thesis sought to examine fatigue experience and CSA in RRMS in everyday life in a case-control study with an ecological momentary assessment (EMA) approach. The primary aims of this doctoral thesis were:

(1) To conduct a detailed prospective investigation comparing fatigue experience in everyday life in pwRRMS to healthy individuals by examining typical diurnal fatigue patterns (changes in fatigue over time within days) and psychosocial contextual effects.

(2) To investigate whether CSA in everyday life differs in pwRRMS compared to healthy individuals, in terms of the CAR, DCS, and the HPA axis-related stress response system (reactivity to daily life stressors)

(3) To explore the relationship between CSA and fatigue, by first conducting a systematic review examining which facets of CSA (e.g., CAR, DCS) are most relevant to fatigue across populations, and then examining these associations prospectively in everyday life in pwRRMS and healthy individuals.

Chapter 2: Methodological considerations and investigation protocol

Each empirical chapter presented in this thesis is a unique analysis of data yielded from a case-control ecological momentary assessment (EMA) study with an RRMS group and healthy control group. Section 2.1 presents the rationale for choosing an EMA strategy. In order to minimise repetition later in the thesis, general information about the protocol, including participant recruitment and characteristics, chosen procedure and measures (salivary cortisol and self-report), and biochemical and statistical analysis are presented in section 2.2. The chapter includes a critical discussion of the use of multilevel modelling (MLM) for EMA data, and concludes by reviewing ethical considerations.

2.1 Introduction to EMA

Several research traditions are principally based on momentary assessment, including *experience sampling* (Csikszentmihalyi & Larson, 1987), *ambulatory assessment* (Ebner-Priemer & Kubiak, 2010), and *EMA* (Stone & Shiffman, 1994). Experience sampling traditionally involves assessing real-time internal-state subjective experience, whereas ambulatory assessment is rooted in physiological and behavioural real-time data capture in everyday life (Ebner-Priemer & Kubiak, 2010). EMA tends to combine both, but contemporary versions of all three traditions converge upon utilizing electronic devices to deliver repeated self-report assessments and/or objective physiological outcomes in everyday life. For simplicity, methods relating to real-time data capture in everyday life will be referred to as EMA in this thesis.

Self-reports are uniquely able to measure internal states including motives, goals, emotions, and perceptions (N. Schwartz, 2008), and involve constructing experience in three ways: "the self in the moment (the experiencing self); the self through time (the remembering self); and the belief-based self (the believing self)" (Conner & Barrett, 2012, p.327). EMA permits the repeated assessment of the experiencing self "in real-time, in real-world settings, over time and across contexts" (Shiffman, Stone, & Hufford,

2008, p.3), capturing experience at that moment (*momentary* assessments) while minimising or eliminating recall biases.

2.1.1 Avoiding recall bias: real-time measurement

Retrospective self-report provision may not emanate from an accurate memory log but instead be the subject of several biasing influences, even upon the initial encoding process (Hufford, 2008). Accurate information retrieval from autobiographic memory (memory of experience) requires the reconstruction of past experience correctly conserved to memory; a process influenced by event-related factors including novelty and salience (Hufford, 2008) and person-related factors such as current mood (P. H. Blaney, 1986). Past unhappy events are recalled more easily while in a negative mood, and positive events are recalled more easily while feeling positive (for example, Clark & Teasdale, 1985): the mood-congruence effect (P. H. Blaney, 1986).

Recall overestimation has been observed in several types of retrospective self-report, including emotional experience (Miron-Shatz, Stone, & Kahneman, 2009) and pain symptoms (Stone, Broderick, Shiffman, & Schwartz, 2004) compared to momentary assessments, and physical activity compared to accelerometer readings (Prince et al., 2008). Recall bias may be influenced by social desirability, with underestimation apparent in socially undesirable behaviours such as alcohol intake in heavy drinkers (Lemmens, Knibbe, & Tan, 1988; Poikolainen, 1985).

2.1.2 Ecological validity: real-world measurement

Rather than acquire summary reports of experience, EMA permits the measurement of ongoing experience in everyday life. The term *ecological validity* typically refers to whether potential effects have been examined in contexts that "accurately represent the typical conditions under which that effect occurs in the real world" (Reis, 2011, p.6). As its name suggests, EMA is principally based on real-world assessment; by studying relationships between variables in everyday life, findings are more generalisable to the real world.

2.1.3 Within-subjects effects: time and context

EMA provides opportunity to observe reliable estimates of both between-subjects and within-subjects variance (J. E. Schwartz & Stone, 1998). As well as enabling symptom measurement at, or near, the point of experience, EMA permits the examination of factors potentially associated with intra-individual symptom variation not possible with retrospective summary measures. Two individuals may have similar average symptom severity over a set period, but where one person experiences the symptom in a fairly consistent manner, another may experience it with periods of low and high severity.

EMA protocols enable observations of thoughts and feelings that are "functionally linked to the underlying neurobiological processes" (Conner & Barrett, 2012, p.332), facilitating real-world observations with synchronous physiological measures such as cortisol (Schlotz, 2011). Examining cortisol levels within saliva in everyday life has become a popular focus of psychoneuroendocrinology research (Kirschbaum & Hellhammer, 1989). It is non-invasive, relatively stress-free, permits frequent sampling (A. Levine et al., 2007), and can be comfortably incorporated into an EMA strategy.

2.1.4 Measurement reactivity

Measurement reactivity refers to the hypothesis that symptom self-monitoring alters the very experience being observed via heightened symptom-focussing (Cruise, Broderick, Porter, Kaell, & Stone, 1996) and is a prominent commonly-raised methodological concern. However, while measurement reactivity makes sense intuitively, there is little empirical evidence. Stone et al. (2003) randomly assigned 91 participants with chronic pain to varying densities of momentary pain assessments (none, 3/day, 6/day, 12/day) for 2 weeks, testing the hypothesis that more frequent assessments increases daily pain scores. Momentary pain was assessed with a 0-100 visual analogue scale (VAS). The study found no evidence of measurement reactivity, as (1) increasing densities of momentary assessments were not associated with elevations in perceived pain ratings over time; and (2) there were no differences in 1-week retrospective pain ratings between those providing EMA (any density) and those who did not.

Measurement reactivity was not evident in other studies of pain in clinical groups (Aaron, Turner, Mancl, Brister, & Sawchuk, 2005; von Baeyer, 1994) or in alcohol consumption in problem drinkers (Hufford, Shields, Shiffman, Paty, & Balabanis, 2002). Furthermore, in a clinical burnout population, measurement reactivity (operationalised by change in ratings over time) accounted for only 0.2% of momentary fatigue rating variability over 2 weeks (Sonnenschein, Sorbi, van Doornen, & Maas, 2006).

2.1.5 Design considerations

EMA aims to investigate daily experience by acquiring momentary assessments forming a representative sample of daily experience, or in the case of cortisol, a representative sample of the cortisol circadian rhythm (Kudielka, Gierens, Hellhammer, Wüst, & Schlotz, 2012; Schlotz, 2011; Stone & Shiffman, 2002). Importantly, the desire for frequent assessments must be balanced against participant burden; everyday life research may not achieve ecological validity if the study is intrusive to the point of distorting its representativeness (Reis, 2011).

An EMA strategy can employ either an event-related or time-based design (or, in some cases, both) to compose assessment schedules within days. An event-related design begins assessments upon event occurrence, and is most useful in observing relatively infrequent discrete events, such as migraine. If the aim is rather to observe continuous experience, such as mood, a time-based strategy should be preferred prompting assessments at regular intervals (Shiffman et al., 2008). An event-related design is utilized to observe the cortisol awakening response (CAR), as the awakening "event" may vary in time between participants and days. Time-based designs may be implemented where the primary interest is observing the cortisol circadian rhythm or daily experience. Ideally, where within-person and between-person variability is explored, the time-based design follows a variable-occasion design where assessments are not fixed to a specific time (as in fixed-occasion designs) but, instead, are quasi-randomly prompted within strategic time windows (Schlotz, 2011). Variable-occasion designs enable representative sampling of daily life and minimise signal anticipation while avoiding sources of systematic bias. For example, participants may manipulate daily routines to establish periods of low demand within which to respond to an anticipated assessment; and fixed-

occasion assessments may give undue influence to situational factors, such as assessments made at 1700h coinciding with leaving the workplace.

Salivary cortisol within-day sampling frequency has varied in previous studies in everyday life. Recent reviews recommend obtaining at least three assessments per day (to adequately capture diurnal slope) over at least 2 days (due to variability of hormone secretion between days) (Kudielka et al., 2012; Schlotz, 2011). These suggestions reflect the physiological nature of cortisol secretion as pulsatile hormone (E. A. Young et al., 2004); if a sample coincides with a cortisol secretion "pulse", statistical model parameters estimated for days with two samples or less would be substantially biased. If intending to examine within-subjects effects, the statistical power to detect such effects increases with more assessments per day and more assessment days per participant; however, participant burden and study cost must still be considered (Kudielka et al., 2012).

2.2 Investigation protocol

This section presents the investigation protocol for the clinical study providing data analysed within this thesis, building upon rationale provided in section 2.1. Given the intention to examine relationships between fatigue, psychosocial factors, and cortisol both within-subjects and between-subjects, momentary self-report and salivary cortisol sampling were incorporated into an EMA design.

The EMA investigation protocol was guided by a handheld electronic device (Hewlett Packard iPAQ 111 Classic Handheld*, Bracknell, UK*) prompting assessments at strategic time-points by auditory alarm, presenting self-report measures via touch-sensitive screen, and allocating time-stamps to responses. The compact handheld had a 3.5inch (88.9mm) touchscreen operated by stylus, and used the Microsoft Windows Mobile 6.0 operating system. The handhelds were programmed by a research technician at the University of Southampton using Microsoft Visual Studio 2010. Battery life was sufficient to last the length of the EMA protocol without recharge.

2.2.1 Participants

This case-control study had an RRMS group and a healthy control group. Recruitment took place between February 2012 and February 2013. This section describes the eligibility criteria for the study, the rationale behind the sample size chosen, the recruitment procedures that were carried out, and finally the size and demographics of the sample recruited.

2.2.1.1 RRMS group eligibility criteria and rationale

The RRMS group inclusion criteria were: (1) a neurologist-confirmed RRMS diagnosis; (2) aged between 18 and 65 years (of working age); and (3) a fluent grasp of the English language. Exclusion criteria were: (1) a relapse or corticosteroid treatment in the last 3 months; (2) low mobility, defined as reporting inability to walk 300 metres without rest and without aid; (3) a diagnosed acute or chronic comorbidity, other than depression which was permitted; (4) pregnancy; and (5) shift work. The definition of "low mobility" in exclusion criterion 2 was changed in June 2012 to "…inability to walk 300 metres without rest, *with* or without aid". During clinics it became evident that many of those excluded on criterion 2 before June 2012 frequently demonstrated equal levels of ambulation to those who were eligible but were using the aid as a precautionary measure.

Inclusion criterion 1 ensured a homogeneous MS sample was recruited, which was important as fatigue prevalence (for example, Mills & Young, 2010) and HPA axis activity (for example, Ysrraelit et al., 2008) have both been shown to differ between MS types. A fluent grasp of English language was important as the handheld had only an English language presentation. Exclusion criterion 1 ensured all RRMS participants were within remission-phase, controlling for the various inflammatory effects of relapse (Sospedra & Martin, 2005) shown to prompt elevated CSA in pwRRMS (Ysrraelit et al., 2008). Exogenous glucocorticoids are the treatment of choice for acute relapses (Milligan et al., 1987) and would disrupt measurement of endogenous cortisol. A 3-month period replicated the criterion chosen previously (Gold et al., 2010). Other disease-modifying therapy (DMT) regimens were permitted and recorded, and were accounted for in relevant statistical analyses. Participant medication profiles are detailed in Appendix C. Exclusion criterion 2 ensured all participants were physically able to carry out the EMA protocol and

controlled for mobility levels between groups, which may impact on physiology. Exclusion criterion 3 ensured observations were relevant to the RRMS population of interest and not due to another disease or condition. In addition, some relatively common conditions such as asthma often involve steroid treatments which impact upon HPA axis activity (Dluhy, 1998). Depression was permitted due to its high prevalence in MS (Siegert & Abernethy, 2005) and its prominent role in recent examinations of CSA in RRMS (Gold et al., 2010; Gold et al., 2011; Kern et al., 2011). HPA axis function can be altered by pregnancy (Carr, Parker, Madden, MacDonald, & Porter, 1981; Jung et al., 2011), and shift work has been associated with altered CSA as workers adapt their circadian systems (Griefahn & Robens, 2010; Horowitz, Cade, Wolfe, & Czeisler, 2001) and is an exacerbating factor for fatigue (Shen et al., 2006).

2.2.1.2 Healthy control group eligibility criteria and rationale

The healthy control group inclusion criteria were: (1) aged between 18 and 65 years; and (2) fluency in the English language. Exclusion criteria were (1) any acute or chronic disease or illness; (2) current use of prescribed medications; (3) primary caregivers; (4) pregnancy; and (5) shift-workers.

The first two exclusion criteria ensured the comparison group were in good health at the time of the study and not ingesting medications that may influence HPA axis activity. Caregiving is often used as a model of chronic stress in psychophysiology studies (for example, Kiecolt-Glaser et al., 1987) so would introduce bias. Pregnant women and shift-workers were excluded as per the RRMS group rationale.

2.2.1.3 Sample size rationale

At the time of the PhD proposal, only one study had been published examining CSA in RRMS in everyday life (Gold et al., 2010), and had not reported sufficient data to compute effect sizes. Data from Ysrraelit et al. (2008) examining 24h urinary cortisol levels in remission-phase RRMS and controls had demonstrated a medium effect size (d = 0.66). The same paper also observed 0800h serum cortisol differences (d = 0.72). Taking the lower effect size, a sample size calculation with good statistical power ($1 - \beta = .80$) based on an independent samples t-test suggested a group sample size of 38 was needed to detect the same effect, if present here. Therefore the primary

recruitment objective was to recruit 40 participants to the RRMS group and 40 age (±3 years) and sex-matched healthy individuals to the control group.

It should be noted that power calculations are not straightforward with multilevel data, and rely upon the number of participants, assessment days, and assessments, as well as intra-class correlation coefficients and quantity of missing data that are not available during study set-up. As such, there is debate as to the feasibility of *a priori* power estimations in multilevel designs, particularly when there are three levels of data (de Jong, Moerbeek, & van der Leeden, 2009). With increasing numbers of days and assessments comes improved precision of parameter estimates in MLM and an increase in power. The EMA design contained 4 days containing nine assessments each, which represents the most robust examination of CSA in RRMS to date. Until very recently, the targeted sample size was larger at every level of data than any other study published examining CSA in RRMS in everyday life (see Table 1).

Table 1 Sample Sizes in Studies examining Cortisol Secretory Activity in Everyday Life in RRMS and Healthy Control groups

	Level – 1 (Assessments)	Level-2 (Days)	Level-3 (Participants)
Gold et al. (2010)	3 (CAR only)	2	45 (29 RRMS, 16 control[a])
Kern et al. (2011)	6 (4 CAR, 2 daytime)	2	48 (32 RRMS, 16 control)
Kern et al. (2013)	6 (4 CAR, 2 daytime)	2	89 (55 RRMS, 34 control)
PhD study design (primary aim)	9 (3 CAR, 6 daytime)	4	80 (40 RRMS, 40 control)

Note. CAR computations aggregated to the Day-level in thesis; CAR and daytime cortisol aggregated to participant level in all other studies.

[a] 20 controls, but 4 excluded due to missing data or unreturned samples.

Major depressive disorder (MDD) is a prevalent comorbidity in MS and has been shown to have an impact on CSA (Knorr et al., 2010); previous studies have suggested this comorbidity is driving greater CSA in RRMS (Gold et al., 2010; Gold et al., 2011; Kern et al., 2011). Therefore, there was a secondary aim to recruit an additional 10 pwRRMS with MDD indicators to perform meaningful sensitivity analyses. However, recruitment stopped once the primary recruitment objective was achieved. The prevalence of pwRRMS and MDD recruited was less than the expected 20% of the population (Siegert &

Abernethy, 2005) and only 6 were recruited, which was insufficient for a sensitivity analysis. Therefore, these individuals were excluded from the final sample. The lower prevalence rate was possibly due to self-selection bias, with those with MDD less likely to choose to take part in the study.

2.2.1.4 Recruitment strategies

2.2.1.4.1 RRMS group

RRMS group participants were recruited from several sources: (1) MS nurse clinics at University Hospital Southampton NHS Foundation Trust (n = 30); (2) neurologist clinics at Guys and St Thomas' NHS Foundation Trust, London (n = 12); (3) UK MS Society website, newsletters, and branch meetings (n = 2); and (4) local (Hampshire) newspaper letters and advertisements (n = 3). In NHS services, MS nurses or neurologists asked consecutive RRMS patients in routine clinics (who they believed met the eligibility criteria) whether they were interested in taking part in the study. Interested individuals then spoke to the researcher about the study after their clinic appointment had ended. A brief yet detailed verbal summary of the study's aims and protocol was given and, without yet committing to taking part, patients left with a participant information sheet (PIS; See Appendix D) to read and consider. Both the recruitment process and the PIS avoided the terms "stress" and "fatigue" to minimise self-monitoring prior to EMA commencement. The health professionals were often unable to check all eligibility criteria, so these were checked by the researcher in the follow-up email and telephone call at least 48 hours later.

The study was advertised by the MS Society on their website *(http://www.mssociety.org.uk/ms-research/get-involved-research/be-in-a-study)*, and local branches informed members via newsletters and email circulars. This route was not very successful as the majority of individuals engaging with the MS Society and contacting the researcher had a progressive MS type or were not ambulatory and so did not meet eligibility criteria. A standard letter was written to local newspapers in Hampshire and surrounding areas, and all published the letter within their letters section. Initial contact was made by the individual and, after checking eligibility criteria, the PIS was sent to the individual for consideration.

2.2.1.4.2 Healthy control group

The healthy control group were recruited from different sources: (1) staff members at the University of Southampton (*n* = 20); and (2) the local community in Hampshire (*n* = 13) and Greater London (*n* = 7). University of Southampton staff members were recruited using a notice placed on the intranet seeking individuals aged within the target demographic for matching to the RRMS group (30 – 50 years). Two posts were placed several months apart. Posters were also placed around the University of Southampton campus. The target demographic was sought from the community through opportunity sampling. The PIS was sent to the individual upon initial communication with the researcher.

2.2.1.5 Recruited sample characteristics

The study recruited 42 pwRRMS and 40 healthy individuals. Of the pwRRMS approached or approaching the researcher to take part, 40% (82/205) met all eligibility criteria, of which 59% (48/82) eventually took part in the study. Note that 6 of 48 pwRRMS had indicators of MDD and were later excluded, as explained in section 2.2.1.3.

There was some redundancy in both groups. In the RRMS group, one participant withdrew after the introductory session citing difficulties participating alongside their job; another withdrew due to a comorbidity diagnosed midway through the EMA protocol; and data was lost from two other participants due to technical faults with handhelds. Two of these redundancies occurred in February 2013 and it was too late to rerun or recruit new participants. In the control group, two exclusions occurred once the saliva samples were analysed due to consistently and excessively high, outlying, cortisol levels, indicative of an endocrine abnormality. Therefore, the final sample contained 38 individuals in the RRMS group and 38 in the control group. Sample characteristics for both groups are presented in Table 2.

Table 2 Demographic and Clinical Characteristics of Study Participants

	RRMS	Control	*p*
n	38	38	
Age	41.89 (7.53)	40.34 (8.16)	.391
Gender (Male : Female ratio)	7:31	7:31	
EDSS	4.35 (1.40)	-	
MS Duration (years)	6.03 (5.18)	-	
Employment			
Paid employment	30	33	
Unpaid employment	3	1	
Unemployed	5	4	
Ethnicity			
White British or White Other	38	35	
Mixed British or Mixed Other	0	1	
Asian	0	2	
Relationship status			
Married or cohabiting	30	28	
Partner, but not cohabiting	2	4	
Single	6	6	
Education Level			
GCSE/O-Level	9	6	
A-Levels	11	5	
Graduate	17	24	
Doctorate	1	3	
Smokers (Yes : No ratio)	4:34	0:38	
Fatigue			
FS Total [a]	17.58 (7.09)	11.55 (2.87)	<.001
FS Physical subscale [a]	11.18 (4.89)	7.26 (2.34)	<.001
FS Mental subscale [a]	6.39 (2.66)	4.29 (0.96)	<.001
MFIS [a]	9.39 (4.25)	3.87 (3.82)	<.001
Stress			
CSSS [a]	19.82 (9.36)	14.11 (7.93)	.006
PSS	16.84 (6.37)	13.16 (6.53)	.015
Anxiety and Depression			
HADS-Anxiety [a]	7.50 (3.90)	4.82 (3.12)	.003
HADS-Depression [a]	4.00 (2.29)	2.08 (2.27)	<.001

Note. Mean (*SD*) shown for all continuous variables. EDSS indicated Expanded Disability Status Scale; FS, Fatigue Scale; MFIS, Modified Fatigue Impact Scale; CSSS, Chronic Stress Screening Scale; PSS, Perceived Stress Scale; HADS, Hospital Anxiety and Depression Scale. Details of scales presented in section 2.2.3.1.

[a] non-parametric Mann Whitney U tests conducted; Otherwise, independent samples t-test.

2.2.2 Procedure

The prospective research design had three phases completed by all participants: (1) an introductory session held at the University of Southampton or Kings College London; (2) 4 consecutive weekdays of EMA incorporating momentary self-reports and saliva assessments using Cortisol Salivettes (Sarstedt*, Leicester, UK*) and led by handheld device; and (3) a 6-month follow-up consisting of self-report questionnaires.

2.2.2.1 Introductory session

Each participant attended a one-to-one introductory session with the researcher, providing written informed consent upon arrival. Participants completed a series of self-report questionnaires (see section 2.2.3.1) via an online portal (*www.isurvey.soton. ac.uk/2310)* and were talked through a demonstration version of the EMA protocol on the handheld. Instructions on how to provide a saliva sample were also given.

Once confident in the EMA protocol, participants left with the following items contained within a large re-sealable bag: (1) four smaller re-sealable bags (Days 1 to 4) containing nine salivettes each; (2) a pre-programmed handheld device; (3) a participant instruction booklet (Appendix E); and (4) a waterproof pen to write on salivette labels.

2.2.2.2 EMA design

This section describes the EMA protocol design, and a critical discussion of the momentary assessments themselves is provided in section 2.2.3.2. EMA was carried out over 4 consecutive weekdays, controlling for weekend-weekday effects in CSA (Kunz-Ebrecht, Kirschbaum, Marmot, & Steptoe, 2004; Schlotz et al., 2004). EMA began on a Monday or Tuesday within 1 week of the Introductory Session. Each day's EMA design was identical, and had 10 assessment events within three distinct component designs: (1) morning CAR assessment design (three events); (2) daytime cortisol and experience design (six events); and (3) end-of-day recall of fatigue experience (one event).

All events were directed by handheld device auditory alarm prompts. Upon each alarm, the device screen lit up and presented a "Start" box which the participant pressed to respond and begin (see Appendix F for an image of the device screen upon alarm). Alarms sounded twice; if the participant did not

respond to the first alarm lasting 30 seconds, a louder alarm occurred 5 minutes later. If the participant did not press “Start” on the handheld screen within 5 minutes of the second alarm, a missed event was recorded and the handheld entered “sleep” mode until the next scheduled event.

With the exception of the end-of-day event, the handheld requested a saliva sample be provided immediately upon each alarm and then presented a series of items requiring a response. The handheld provided an accurate time-stamp for each saliva sample and item response. Participants were instructed to respond to all events and to follow all presented instructions. An overview of the items presented at each event is detailed in Figure 1 (morning protocol) and Figure 2 (daytime and end-of-day protocol).

2.2.2.2.1 Morning event-related fixed design

The CAR was assessed by the first three samples of the day (S1–S3): upon awakening (S1/T0); 30 minutes after awakening (S2/T30); and 45 minutes after awakening (S3/T45). Three assessments, which must include a waking sample, are considered adequate to reliably estimate the CAR using area under the curve estimates (Clow, Thorn, Evans, & Hucklebridge, 2004; J. C. Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). The “event” in this event-based design is awakening. The S1 awakening event was initiated in one of three ways: (1) pre-set auditory alarm, similar to an alarm clock; (2) automatic auditory alarm at 0830h if not already awake; or (3) manual initiation upon awakening by pressing the “Start” box (available from 0400h). If the S1 alarm was pre-set, but awakening occurred earlier than expected, participants were still able to manually initiate the handheld. The automatic alarm at 0830h restricted the awakening time range as awakening times can affect cortisol levels (Edwards et al., 2001), and also prevented CAR design overlap with the later 1000h – 2000h design.

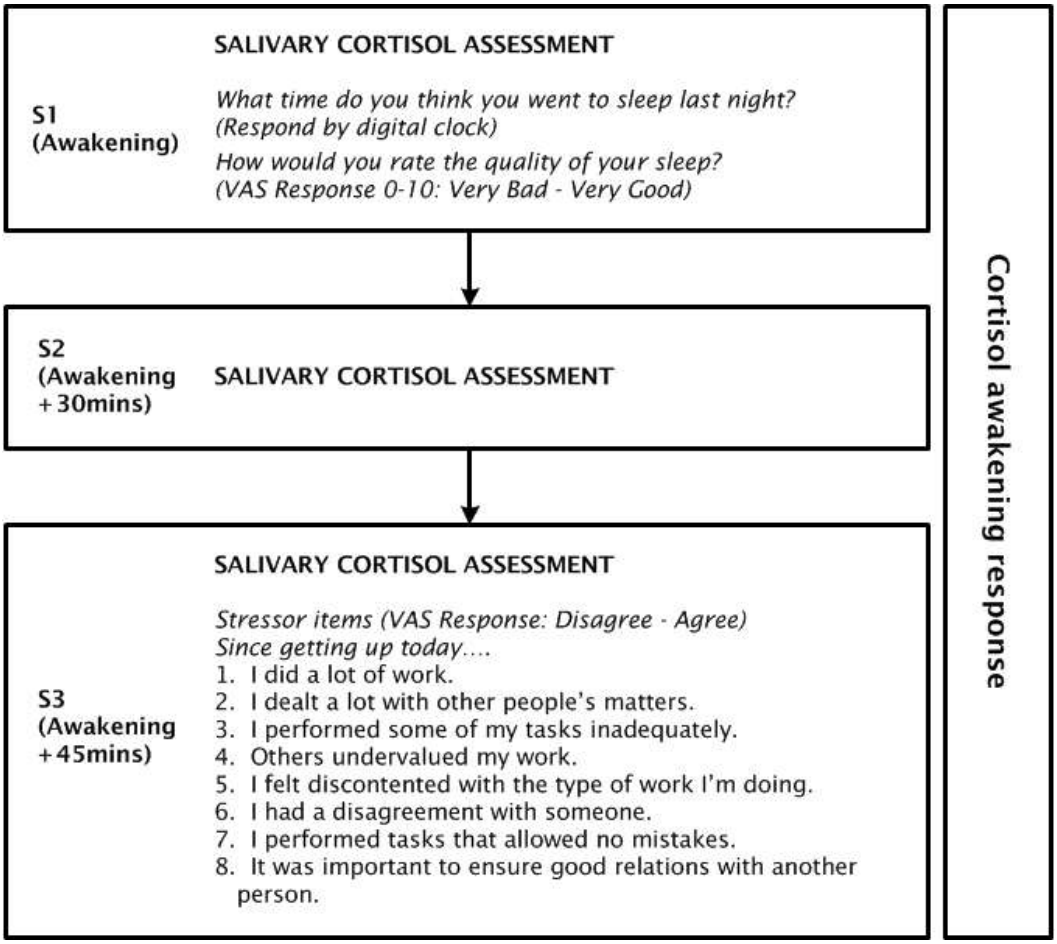


Figure 1. Assessment design for cortisol awakening response.

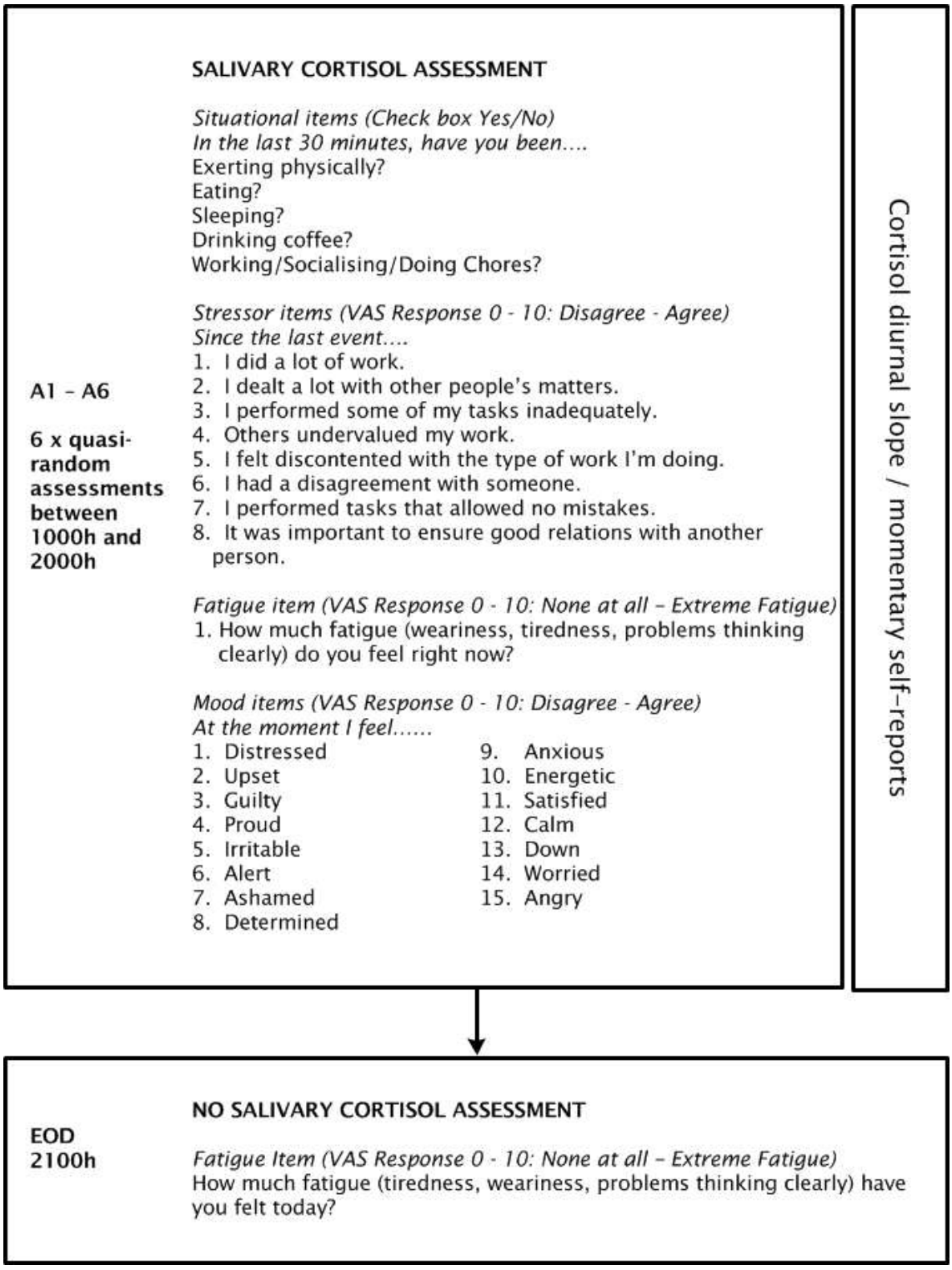


Figure 2. Assessment design for diurnal cortisol slope, diurnal fatigue pattern, daily psychosocial experience, and end-of-day recall.

2.2.2.2.2 Daytime time-based variable-occasion design

The diurnal cortisol slope (DCS) and diurnal fatigue patterns were modelled using six quasi-random momentary assessments (referred to henceforth as A1 to A6). A1 to A6 occurred between 1000h and 2000h under constraints of no two events <30min or >100min apart. This design enabled representative sampling of daily life experience and cortisol output over this period, and permitted within-subjects examinations (Schlotz, 2011).

2.2.2.2.3 End-of-day fixed-occasion design

At 2100h each day, an alarm prompted the end-of-day event. Participants were presented with items asking them to recall fatigue experience over the preceding day. Once the end-of-day event was completed, the handheld asked participants to store used salivettes in a refrigerator and to keep handhelds close-by when they went to bed. The handheld then entered "sleep mode" until the following morning.

2.2.2.2.4 Maximising protocol compliance

EMA validity can be compromised by participant noncompliance (Stone & Shiffman, 2002), and maximising adherence to a sampling protocol is important in salivary cortisol studies in particular as hormones vary quickly over time (Broderick et al., 2004). For example, Kudielka, Broderick, and Kirschbaum (2003) monitored compliance to a one-day saliva sampling protocol electronically using drug exposure monitors (*Aardex*, Switzerland), finding the peak minus nadir CAR was far greater in those who were compliant than those who were not (11.03 vs. 3.31 nmol/L).

In the present study, audible prompts minimised the reliance on participants remembering assessment schedules. In order to maximise compliance, a strategy was adopted first used by Stetler, Dickerson, and Miller (2004): upon the alarm, the handheld presented a random 3-digit code that participants were instructed to transfer to the label of the salivette being used. Stetler et al. (2004) acquired 96.5% usable samples (samples with correct codes) from the total requested, and there was similar compliance (93.0%) in another study (Powell & Schlotz, 2012). Samples with missing or incorrect codes were removed.

Measuring cortisol immediately upon awakening is paramount to CAR studies due to the speed of the cortisol response to waking (Clow et al., 2004; Kudielka et al., 2003). To minimise delay or the chance of forgetting, participants were able to initiate the S1 awakening event in one of three ways described previously in section 2.2.2.2.1. Being woken by auditory alarm, rather than spontaneously, is not thought to influence the CAR (Wüst et al., 2000). It should be noted that time of awakening in the present study was indicated by the S1 morning alarm or manual initiation; it was not possible to ascertain when participants actually woke.

2.2.2.2.5 Handheld device usability functions

Participants were able to engage with the handheld between events to use several incorporated usability functions which sought to minimise participant burden and interference with daily activities. For the A1 to A6 events and end-of-day event, participants could postpone the event upon hearing the alarm for 5, 10, or 15 minutes if needed. The S1 to S3 events could not be postponed due to the fixed event-related design and, as stated, delays in sampling can invalidate the CAR.

Participants could choose to enter "Silent Mode" via a *Help* option on the handheld homepage, which prevented an auditory alarm. Participants were asked to use this option sparingly. Silent Mode could be entered for a predetermined period of time (15, 30, 45, or 60min) or for the next event only. After one silent event, handhelds automatically reverted back to Loud Mode to prevent participants forgetting about the device; however, participants were able to select Silent Mode again immediately if needed. Silent Mode consisted of the handheld screen lighting up and producing three very quiet "click" sounds. Participants were asked to have the handheld within view (out of its carry-case) if in Silent Mode. There was no inbuilt vibrate function.

2.2.2.2.6 Return of samples and handhelds

Upon EMA protocol completion, participants returned to the laboratory at the University of Southampton or Kings College London with their handheld and used salivettes. There they were thanked and debriefed about the study so far. Participants were reminded they would be contacted again in 6 months to complete follow-up questionnaires. EMA data from the handheld was

immediately downloaded to the laboratory PC and processed using Matlab Version 7.10.0.

2.2.2.3 Follow-up

A 6-month follow-up provided data to be compared to baseline in the longitudinal analysis presented in Chapter 6. Participants were sent a questionnaire pack in the post 6 months after completion of the EMA protocol. Self-reported fatigue, perceived stress, and employment and health status changes were measured, and neurological disability was assessed in the RRMS group only. Details of these scales are presented below. Questionnaire packs took around 10–15 minutes to complete, and were sent in paper format as not all questionnaires could be presented online. Return was by pre-paid mail.

2.2.3 Measures

In this section, baseline questionnaires presented within the introductory session are detailed first, along with details of validation and reliability. With the exception of the neurological disability questionnaire (see below), baseline questionnaires were completed by participants in both groups, and all were presented via an online portal. Some baseline questionnaires were repeated at follow-up as indicated below. An overview of the momentary assessments presented within the EMA protocol follows.

2.2.3.1 Baseline questionnaires

2.2.3.1.1 Neurological disability

Disability was assessed in the RRMS group by the self-administered Expanded Disability Status Scale (EDSS; Bowen, Gibbons, Gianas, & Kraft, 2001) which is MS-specific and covers a spectrum of functioning including mobility, visual disturbance, motor co-ordination, strength, cognitive impairment, and bladder dysfunction. The EDSS was presented at baseline and follow-up. Participants are asked to consider their symptoms “on an average day, at your best”. Items include “I can walk 500 metres without rest” with the response checkboxes “Yes” or “No”. Further information is gained from the sub-item “If yes, I can do this…” with responses “…without help”, “…with a cane”, “…with two canes”, “…with a walker”. Items regarding strength, co-ordination, sensation, bowel function, bladder function, vision (glasses permitted),

diplopia (double vision), unsteady eyes, speech, swallowing, and cognition follow, and have their own bespoke response options. The EDSS has a bespoke rating system for each item, contributing to an overall score ranging from 0.0 (no neurological impairment) to 10.0 (death from MS).

During development, the self-administered EDSS was shown to be highly correlated (r = .89) with the original physician-delivered EDSS (Kurtzke, 1983) considered the "gold standard" clinical measure for MS disability status (Goldman, Motl, & Rudick, 2010; Rudick et al., 2010). This correlation was similar to the inter-rater reliability of the physician-administered test (r = .87; Bowen et al., 2001). The measure is advantageous in not needing neurologist time, and has been incorporated into several studies in MS (for example, Chwastiak et al., 2005; Dennison, Moss-Morris, Silber, Galea, & Chalder, 2010). The EDSS was the only baseline measure not presented online as its unique response format was incompatible, so was completed by hand.

2.2.3.1.2 Socio-demographic and general health questionnaire

Self-reports were provided on age, sex, ethnic origin, relationship status, employment status, smoking status, menstrual phase and oral contraceptive use in female participants, and current medication regimen. Different menstrual phases (follicular, days 1-14; and luteal, days 15-28) and oral contraceptive use have been shown to exert effects on HPA axis activity (Altemus et al., 1997; Kirschbaum, Kudielka, Gaab, Schommer, & Hellhammer, 1999), and smoking has a stimulatory effect on the HPA axis (Rohleder & Kirschbaum, 2006).

2.2.3.1.3 Fatigue

Fatigue was measured at baseline using the Fatigue Scale (FS; Chalder et al., 1993) and the Modified Fatigue Impact Scale (MFIS-5; Ritvo et al., 1997). The two scales measure different facets of fatigue experience in MS: symptom severity and its impact on daily life, respectively. Both were completed at baseline and follow-up.

The 11-item FS (overall α = .89) measures fatigue severity over the last month and has two subscales examining physical fatigue (7 items, e.g., "Do you have problems with tiredness?" α = .85) and mental fatigue (4 items, e.g., "Do you have difficulty concentrating?" α = .82). Responses are on a 4-point

Likert scale (better than usual, no more than usual, worse than usual, much worse than usual). The FS has been validated in many populations including chronic fatigue syndrome (Morriss, Wearden, & Mullis, 1998), has been used in several MS studies (for example, Skerrett & Moss-Morris, 2006), and has received a good rating for usability (easy to understand/complete, minimal burden) in a review of fatigue measures in chronic illness (Whitehead, 2009). The FS is scored on a continuous scale (0 – 33), and can also be used to distinguish between those with and without "clinically meaningful fatigue" using a bimodal scoring system (Chalder et al., 1993). In the bimodal scoring system, items rated 0 or 1 are scored 0, and those rated 2 or 3 are scored 1. Individuals scoring ≥4 on the bimodal system are categorised as having "clinically meaningful fatigue". This cut off was shown to have reasonable sensitivity (75.5) and specificity (74.5) compared to the fatigue item of the Clinical Interview Schedule (Lewis, Pelosi, Araya, & Dunn, 1992).

The MFIS-5 is an abbreviated 5-item version of the Fatigue Impact Scale (Fisk, Ritvo, et al., 1994) and represents the 5 items correlating most strongly with the total score on the original 21-item scale. The MFIS-5 (α = .82) assesses fatigue's impact on daily life over the prior 4 weeks and was initially developed and validated within an MS population. The MFIS-5 includes items such as "Because of fatigue, I have been limited in my ability to do things away from home". Response is via a 5-point Likert scale (never, rarely, sometimes, often, almost always) and MFIS-5 score is the sum of item scores (range: 0-20). The MFIS-5 is a component of the MS Quality of Life Inventory (Ritvo et al., 1997) and has been recommended for use within the MS population by the Multiple Sclerosis Council for Clinical Practice Guidelines (1998).

2.2.3.1.4 Stress

As the HPA axis is highly sensitive to psychosocial stress (Dickerson & Kemeny, 2004), it was important to measure baseline stressful experience effectively. Two measures assessing different aspects of stress were administered: (1) the Chronic Stress Screening Scale (CSSS; Schulz, Schlotz, & Becker, 2004); and (2) the Perceived Stress Scale (PSS; S. Cohen et al., 1983). Both were completed at baseline and follow-up.

The CSSS measures chronically-experienced stress over the prior 3 months, and is a 12-item scale derived from items within the English language

version of the Trier Inventory of Chronic Stress (Schulz et al., 2004). Chronic stress has been associated with many disorders, and HPA axis dysfunction may be a mechanism underlying this relationship (Heim et al., 2000). In contrast to acute stress, chronic stress describes long-lasting demands, or demands recurring with high frequency. The CSSS defines an individual as chronically stressed if they are worrying a lot, feeling overextended and overwhelmed, and receiving no recognition for their efforts. Items include "Times when my worries overwhelm me", and responses are via 5-point Likert scales (never, rarely, sometimes, often, very often). Scores range from 0-48 and high scores indicate high levels of chronic stress. The scale demonstrated excellent internal consistency in healthy individuals (α = .91) (Petrowski, Paul, Albani, & Brahler, 2012).

The PSS is a 10-item questionnaire based on the theory of stress appraisal (Lazarus & Folkman, 1984), assessing the extent to which individuals appraise situations as stressful, with high scores indicative of high levels of perceived stress. The PSS includes items such as "In the past month, how often have you been able to control the irritations in your life?" and responses are on 5-point Likert scales (never, rarely, sometimes, often, very often), with scores ranging from 0-40. The PSS is the most widely-used instrument for measuring perceptions of stress, and demonstrated excellent internal consistency in an MS population (α = .91) (Wu & Amtmann, 2013) and good or excellent internal consistency in three separate healthy groups (all α > .80) (S. Cohen et al., 1983). The PSS has been incorporated into previous studies investigating the effects of stressful experience in MS (for example, Ackerman, Martino, Heyman, Moyna, & Rabin, 1998; Trojan et al., 2007).

2.2.3.1.5 Depression and Anxiety

Levels of depression within MS are relatively high (Siegert & Abernethy, 2005), and both anxiety and depression are often associated with elevated cortisol levels (Kurina, Schneider, & Waite, 2004; Veen et al., 2011). Anxiety and depression were measured by the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983). The HADS consists of 14 items with two 7-item subscales for anxiety (HADS-A) and depression (HADS-D). Each item of the HADS has its own unique 4-point Likert scale (scored 0 – 3); for example, the item "I can sit at ease and feel relaxed" is responded to with "definitely", "usually", "not often", or "not at all". The HADS has been widely used within MS

research (for example, Janssens et al., 2003; Roosendaal et al., 2009; Wood et al., 2012). High scores reflect high levels of depression or anxiety on their respective subscales (maximum score of 21).

Importantly, the HADS does not include any somatic symptoms of depression or anxiety, such as fatigue or dizziness. There is some debate as to whether to include or exclude items about somatic symptoms in other measures such as the Beck Depression Inventory, including within an MS population (Aikens et al., 1999; Mohr et al., 1997). The HADS has been used in many empirical studies, with a systematic review of its psychometric properties finding support for the two-factor solution, "good" to "very good" concurrent validity with other (longer) measures of anxiety and depression, and the vast majority of constituent studies found Cronbach $\alpha > .75$ for each subscale (Bjelland, Dahl, Haug, & Neckelmann, 2002).

2.2.3.2 Momentary assessments

Saliva samples for free cortisol levels were requested at events S1 to S3 to measure the CAR and at every quasi-random event (A1-A6) to measure the DCS. Time of awakening was recorded as S1 timestamp. At events A1 to A6, participants were presented with several items pertaining to daily life stressors, activities, fatigue, and mood (see Figure 2). There were also self-reports requested during the fixed event morning design (see Figure 1).

2.2.3.2.1 Salivary cortisol

Nine salivary cortisol assessments were made per day (S1–S3; A1–A6), using Cortisol Salivettes (Sarstedt, *Leicester, UK*): plastic vials containing a polyethylene (synthetic) swab which absorbs saliva when held within the mouth and gently chewed for approximately 30 seconds. Cortisol levels are reliably measured using this non-invasive saliva swab procedure (Kirschbaum & Hellhammer, 1989). Cortisol within saliva has been shown to be strongly associated with blood serum levels ($r = .85$; Gallagher et al., 2006) but represents a more viable means of measuring cortisol in everyday life (Kirschbaum & Hellhammer, 1989). Polyethylene swabs were preferred to cotton swabs due to cotton production variability introducing random error; for example, in an unpublished investigation of 30 batches of cotton salivettes, different concentrations of cortisol solutions were pipetted onto swabs and, although the within-batch loss in cortisol was relatively constant, there was

large cortisol loss variation between batches, ranging from 5.7% to 44.4% (cited in Kudielka et al., 2012).

During the CAR measurement period (S1-S3) participants were instructed not to eat, drink anything other than water, smoke, brush their teeth, or exercise. These behavioural restrictions reduced the risk of minor oral abrasions introducing blood into the samples (cortisol levels are higher in blood than saliva) and prevented the potentially confounding effects of smoking, exercise, and caffeine. These behavioural restrictions are commonplace in CAR research (Chida & Steptoe, 2009). During the rest of the day (events A1-A6), there were no restrictions on behaviours before saliva sampling but potential confounders (eating, drinking coffee, sleeping, physical exertion) were monitored by self-report.

Three facets of the CAR were measured: (1) the area under the curve with respect to ground (AUCg); (2) the area under the curve with respect to increase (AUCi); and (3) S1 cortisol level. These are standard CAR computations representing the estimated total cortisol output (AUCg), estimated cortisol output above the waking cortisol level (AUCi), and end of the pre-awakening rise in cortisol (S1) (Clow, Hucklebridge, Stalder, Evans, & Thorn, 2010; Fekedulegn et al., 2007; J. C. Pruessner et al., 2003). Both the AUCg and AUCi were computed from constituent samples (S1-S3; formulae presented in Appendix G). The DCS is a standard means of operationalising cortisol output during the day, and was computed as the linear slope parameter in multilevel models of A1-A6 cortisol levels.

2.2.3.2.2 Morning self-report measures

At S1, participants were asked two questions regarding sleep quality and the recalled time they went to sleep (see Figure 1 for item wording). An indicator of hours slept was computed as the time elapsed between self-reported time of sleep and S1 event. Reduced sleep quality has been shown to be a significant predictor of both overall fatigue and physical fatigue in MS (Trojan et al., 2007), and was associated with an attenuated CAR in civil servants (Hansen et al., 2012).

2.2.3.2.3 Daily self-report measures

Repeated momentary assessments were presented at events A1 to A6. Each event was identical and item order was not randomised, which made presentation as straightforward as possible for participants. Items regarding current activities and behaviours, daily life stressors, fatigue, and finally mood were presented (see Figure 2).

2.2.3.2.3.1 Activities and behaviours

Participants were asked about their current activity using a check box response to options "Working/Studying", "Conducting Household Chores", "Socialising", "Commuting", "Shopping", and "Other". "Yes" or "No" responses were then requested to items regarding whether participants had eaten a meal, slept, exerted physically, smoked a cigarette, or drank coffee in the preceding 30 minutes. All of these behaviours can influence HPA axis activity and were measured to include in analyses as potential covariates.

2.2.3.2.3.2 Daily life stressor measurement

Eight stressor items prefixed with "Since the last event…" were presented by VAS (0 = "Not at all" to 10 = "Very much so") with a "not applicable" option resulting in a missing value. "Not applicable" was available as some items assumed engagement in a task, which may not have been the case. Items included "I did a lot of work". Each item was based on one of eight domains of the Trier Inventory of Chronic Stress (Schulz et al., 2004): Work Overload, Social Overload, Excessive Demands at Work, Lack of Social Recognition, Work Discontent, Social Tensions, Pressure to Perform, and Social Isolation.

These stressor items have been used in a previous EMA study (Powell & Schlotz, 2012); however, a new exploratory factor analysis of the items was conducted using the larger dataset from the present study for data reduction purposes. Principal axis factoring iterations, recommended by Costello and Osborne (2005) for data violating normality assumptions was used. All usable data from 1661 completed events were used. Missing values for items where participants responded "not applicable" resulted in the exclusion of 341 events, leaving 1320 events for the FA. The Kaiser-Meyer-Olkin (*KMO*) statistic is a measure of sampling adequacy, where scores should be a minimum of .50, .70 to .80 represents "good" sampling adequacy, .80 to .90 "great", and >.90

"superb" (Kaiser, 1974). Sampling adequacy for the factor analysis was good (*KMO* = .71) and all *KMO* statistics for individual items were above .67. Bartlett's test of sphericity, χ^2 = 1451.4, *df* = 28, *p* < .001, rejected the null hypothesis that all correlations between items were 0, suggesting factor analysis was appropriate.

Visual inspection of the Scree plot (see Appendix H) suggested a two factor solution, and oblique oblimin rotation permitted factors to correlate (Costello & Osborne, 2005); it is likely that different stressor dimensions are related within-subjects. A two-factor solution (Eigenvalues: 2.40; 1.45) meeting the Kaiser criterion (extraction of all factors with eigenvalues above 1) yielded 48.2% of the variance in the daily stressor items. The first factor reflected *Stressor Challenge* and comprised four items ("I dealt a lot with other people's matters"; "I did a lot of work"; "I performed tasks that allowed no mistakes"; "It was important to ensure good relations with another person"). The second factor reflected *Stressor Hindrance* and comprised four items ("Others undervalued my work"; "I performed some of my tasks inadequately"; "I had a disagreement with someone"; "I felt discontented with the type of work I'm doing"). Factor loadings are reported in Table 3. *Stressor Challenge* had good or acceptable internal consistency computed upon the first quasi-random event (A1) of each day (Day 1, α = .65; Day 2, α = .74; Day 3, α = .72; Day 4, α = .74), and *Stressor Hindrance* internal consistencies ranged from relatively low to good (Day 1, α = .66; Day 2, α = .63; Day 3, α = .57; Day 4, α = .61). It should be noted that alphas are likely to be underestimated due to the non-normality of the response variable distributions (Sheng & Sheng, 2012).

Both factors were scored as the mean of all constituent items within a given event. The literature on daily stressors was examined and factor labels came from concepts developed within the job stress literature: job challenges and job hindrances (Van den Broeck, De Cuyper, De Witte, & Vansteenkiste, 2010). The Job Demands-Resources model (Bakker & Demerouti, 2007) proposed that jobs have two characteristics: demands and resources. Demands, which could be intuitively substituted with "Potential Stressors", has been further teased apart by van den Broeck et al. (2010) into (1) challenges, which are effortful but can lead to goal achievement and stimulation; and

(2) hindrances, which are potential barriers to goal achievement including conflict evoking negative feeling.

Daily life stressor items asked for recall of stressors since the last event, encapsulating all daily stressors over the day and meaning that momentary assessments of mood and cortisol/fatigue (at that moment in time) came at the end of the period from which stressors were recalled. In addition to asking about daily life stressors at events A1 to A6, the same items were presented at S3 to capture morning stressors prefixed with "Since getting up today…"

Table 3 Factor Loadings for Daily Stressor Items from an Exploratory Factor Analysis

Item wording	Mean (*SD*)	Factor loadings (pattern matrix) Stressor Challenge	Stressor Hindrance
Since the last event…..			
I dealt a lot with other peoples matters	4.38 (3.52)	.661	.086
I did a lot of work	5.25 (3.12)	.632	-.088
I performed tasks that allowed no mistakes	1.50 (2.11)	.588	-.015
It was important to ensure good relations with another person	5.98 (3.43)	.433	.072
Others undervalued my work	1.29 (2.11)	-.013	.719
I felt discontented with the type of work I'm doing	2.61 (2.96)	.137	.451
I had a disagreement with someone	0.78 (1.86)	-.095	.446
I performed some of my tasks inadequately	1.50 (2.11)	.083	.410

Note. Tinsley and Tinsley (1987) suggest .30 as a minimum loading for any item to be included within a factor; all item loading parameters >.30.

2.2.3.2.3.3 Fatigue

Momentary fatigue (MF) has been measured in a small but growing number of studies (for example, Broderick et al., 2008; Curran, Beacham, & Andrykowski, 2004; J. Kim, Kikuchi, & Yamamoto, 2013) with each using a single item. Single items reduce participant burden in EMA protocols, which

was an important consideration (Shiffman et al., 2008); however, steps must be taken to avoid ambiguity in meaning. Two studies employing MF single-item assessments have clarified "fatigue" in parentheses (Broderick et al., 2008; Curran et al., 2004). Both studies' MF items closely, but not identically, resembled the single-item Brief Fatigue Inventory (BFI; Mendoza et al., 1999) which asks for fatigue ratings on VAS and contains the words "tiredness" and "weariness" within parentheses to conceptualise fatigue. J. Kim et al. (2013) measured MF with the descriptor "Fatigued" and a 0-100 VAS with no clarification of its meaning. Although "fatigue" could be argued to have high face validity, Chapter 1 described a lack of consensus in its definition that could lead to various interpretations. Therefore, a single-item MF measure was used in the present study developed from the BFI, with an amendment made to parentheses to include "problems thinking clearly". This clarified that fatigue was to be conceptualised as including the mental fatigue dimension as well as physical fatigue. This was important given the generally accepted definition of MS fatigue as "a subjective lack of physical and/or mental energy…" (Multiple Sclerosis Council for Clinical Practice Guidelines, 1998, p.2). The item was worded: "How much fatigue (tiredness, weariness, problems thinking clearly) do you feel right now?" Responses were by VAS from 0 ("No Fatigue") to 10 ("Extreme Fatigue").

A primary advantage of momentary assessment is in assessing phenomena directly from experience, bestowing ecological, face, and content validity so long as the measure is transparent in its meaning (Shrout & Lane, 2011). To determine convergent and discriminant validity, associations of MF with momentary ratings of mood items "Energetic" and "Alert" (convergent, expecting strong associations) and "Anxious" and "Distressed" (discriminant, expecting weak associations) were examined in multilevel models of MF (specified in Equation 1, Chapter 3, p.99) with effects permitted to vary between days and individuals. Validity was demonstrated by strong negative associations with both "Energetic" ($\gamma = -0.53$, $p < .001$) and "Alert" ($\gamma = -0.47$, $p < .01$), and weak associations with both "Anxious" ($\gamma = 0.18$, $p = .33$) and "Distressed" ($\gamma = 0.08$, $p = .74$).

2.2.3.2.3.4 Momentary mood measurement

Fifteen mood adjectives, such as “Distressed”, “Worried”, and “Alert” were presented at events A1 to A6 with the statement “At the moment I feel….”, responded to via VAS (0 = “Not at all” to 10 = “Very much so”). For these items there was no “Not Applicable” option. These mood items have been used in a previous EMA study (Powell & Schlotz, 2012). An exploratory factor analysis was conducted of the 15 items using principal axis factoring and oblique oblimin rotation. There were 1661 events, but 115 events were excluded due to a handheld device programming error allowing six participants use of the “not applicable” response to the mood items, leaving 1546 events for the factor analysis. Visual inspection of the scree plot (see Appendix H) led to a two-factor solution (Eigenvalues: 5.77; 2.44) explaining 54.7% of the variance. *KMO* (.88), KMO for individual items (all above .78) and Bartlett’s test of sphericity $\chi^2 = 10949.3$, df = 105, $p < .001$ suggested factor analysis was appropriate. Factor 1 contained ten items (“Upset”; “Distressed”; “Angry”; “Anxious”; “Irritable”; “Down”; “Worried”; “Ashamed”; “Guilty”; “Calm” (reversed)) reflecting *Distress*. Factor 2 had five items (“Alert”; “Energetic”; “Determined”; “Satisfied”; “Proud”) reflecting *Positive Mood.*

Factor loadings are presented in Table 4. Factor cross-loadings (Costello & Osborne, 2005) were apparent for the “Calm” and “Satisfied” items. The “Calm” item was excluded due to having the smallest contribution to Factor 1 and being a somewhat ambiguous term; it could mean the absence of anger or anxiety, but also experiencing quiet surroundings. The “Satisfied” item was retained as it had a relatively large loading on Factor 2, it fitted well with the factor conceptually, and omitting it left Factor 2 with few items. The final *Distress* scale contained nine items and the *Positive Mood* scale contained five items.

Conceptually, the two factors were variants of positive and negative affectivity. A common assumption has been that positive and negative mood exist upon one dimension (bipolarity), meaning two distinct factors would not be intuitive. This assumption implies positive mood induction eliminates negative mood in a mutually exclusive manner. However, literature examining mood at the momentary level has argued that mixed feelings are possible and “an individual with high positive affect (mood) does not necessarily have low negative affect (mood), and vice versa” (Schimmack, 2008, p.115). For

example, Schimmack (2001) demonstrated mixed feelings in response to mood-inducing pictures using single-item momentary measures of “pleasure” and “displeasure”.

Table 4 Factor Loadings for Momentary Mood Items from an Exploratory Factor Analysis

		Factor loadings (pattern matrix)	
	Mean (*SD*)	Distress	Positive Mood
At the moment, I feel…..			
Upset	1.00 (1.98)	.806	.023
Distressed	1.06 (1.95)	.738	.054
Angry	0.94 (1.85)	.731	.040
Anxious	2.85 (2.85)	.680	-.089
Irritable	2.02 (2.68)	.667	-.116
Down	1.74 (2.63)	.659	-.227
Worried	2.33 (2.84)	.625	-.027
Ashamed	0.60 (1.45)	.620	.155
Guilty	0.86 (1.89)	.585	.105
Calm [a]	6.72 (2.67)	-.531	.292
Alert	6.15 (2.72)	-.042	.764
Energetic	5.12 (2.73)	-.106	.744
Determined	5.58 (2.74)	.123	.641
Satisfied [b]	5.80 (2.58)	-.334	.605
Proud	3.68 (2.91)	.055	.520

Note. Tinsley and Tinsley (1987) suggest .30 as a minimum loading for any item to be included within a factor; all item loading parameters above .30.

[a] Double-loading and item excluded. See text for justification.

[b] Double-loading and item retained. See text for justification.

2.2.3.2.4 End-of-day fatigue measurement

In order to perform analyses of peak-and-end heuristic biases in recalled fatigue in RRMS (presented in Chapter 3), the handheld presented

fatigue recall items at 2100h each day. The single MF item was reworded to encapsulate the day as a whole, and again responded to on a VAS from 0 ("No Fatigue") to 10 ("Extreme Fatigue"): "How much fatigue (tiredness, weariness, problems thinking clearly) have you felt today?"

2.2.4 Storage of saliva samples

Salivettes were stored by participants in a refrigerator at the end of each sampling day, as mould can develop in used salivettes within 4-7 days at room temperature (Nicolson, 2008). Refrigerating each sample immediately upon sampling was unnecessary as salivary cortisol is stable at room temperature for up to 4 weeks without significant degradation (Kirschbaum & Hellhammer, 1989; Schlotz, 2011).

Upon completion of the EMA protocol, participants returned salivettes to the laboratory where they were stored together at -20°C on site. Storing saliva samples at -20°C was considered appropriate as a previous study had observed no significant decreases in cortisol when stored at -20°C for 12 months (Garde & Hansen, 2005). In the present study, saliva samples were not stored for longer than 12 months (median storage time = 17.5 weeks; range = 0.5 weeks, 51.5 weeks).

2.2.5 Biochemical analysis

Once all participants completed the EMA protocol, samples were sent in one batch to the Biochemical Laboratory at the Division of Theoretical and Clinical Psychobiology, University of Trier, Germany. Analysis for concentration of free cortisol content (nmol/L) was by time-resolved immunoassay with fluorescent detection (Dressendörfer, Kirschbaum, Rohde, Stahl, & Strasburger, 1992). Transit of saliva samples in a single batch was necessary as each assay has a unique intra-assay and inter-assay coefficient (reliability) and to analyse in separate batches would introduce substantial bias. Each sample was measured in duplicate, with an intra-assay coefficient of variance (CV) between 4.0% and 6.7%, and inter-assay CV between 7.1% and 9.0%. Intra-assay and inter-assay coefficients below 10% are considered good (Schultheiss & Stanton, 2009).

The detection limit (sensitivity) for the assay was 0.173 nmol/L and gives the lowest concentration of salivary cortisol distinguishable from a sample with no cortisol. There is no detection limit "standard" within the field, and it is not commonly reported within study publications, but 0.173 nmol/L compares favourably with other studies reporting detection limits ranging from 0.33-0.50 nmol/L (Clow et al., 2006; Peeters, Nicholson, & Berkhof, 2003; Schlotz, Schulz, Hellhammer, Stone, & Hellhammer, 2006).

2.2.6 Statistical analysis using multilevel modelling

The present study yielded two datasets that were hierarchical, nested, and multilevel: (1) a 3-level dataset with A1-A6 assessments (k = 76*4*6 = 1824; Level-1) nested within assessment days (m = 76*4 = 304; Level-2), nested within participants (n = 76; Level-3); and (2) a 2-level dataset with daily CAR computations (k = 4*76 = 304; Level-1) nested within individuals (m = 76; Level-2). Group (RRMS or control) did not specify a level of data, but rather a binary variable (0, 1) at the highest data level. In this section, a conceptual rationale for the use of MLM in this type of dataset is presented. Details of specific hypothesis-led analyses are contained within respective chapters. All analyses were completed with SPSS Version 20.0.

Using conventional statistical techniques with multilevel data requires disaggregation or aggregation to manipulate data structures to suit the method. Nested data may be handled at the lowest single level of data (disaggregation), ignoring its clustered nature to perform multiple regression (Hox, 1995). However, this method ignores intra-class correlation coefficients, violating a key assumption of most traditional statistical methods: the independence of observations. The attribution of variance is therefore erroneous, with variance at higher-level units (for example, participants) falsely attributed to lower-level units (for example, assessments) (Hox, 1995). Alternatively, aggregation of lower-level data into higher-level data enables ANOVA using person-means, but eliminates within-individual information, ensuring no valid inferences can be made about factors influencing within-person variation or changes over time.

Where data is limited to repeated measures within individuals (two levels), repeated measures ANOVA can negate some of the problems detailed above (Girden, 1992) but the assumptions underlying this technique are strict,

requiring: (1) balanced datasets, with no missing assessments; (2) data collected at fixed times for each participant; and (3) covariates (predictors) that are stable over time (J. E. Schwartz & Stone, 2008). These assumptions are rarely achieved in EMA studies, and were not here for the following reasons: (1) missing data was apparent (discussed further later); (2) a variable-occasion design was employed in part (A1-A6 events), and (3) some predictors were measured at the momentary level and varied over time.

MLM has emerged as a flexible statistical tool for the analyses of data with hierarchical and nested structures (Nezlek, 2007; Snijders & Bosker, 2012). A variety of terms for similar analyses exist within the literature, including *mixed-effects modelling* (Laird & Ware, 1982), *hierarchical linear models* (Raudenbush & Bryk, 1986), and *random regression models* (Gibbons et al., 1993). However, each can be considered a series of regression equations within a hierarchical system (Hox, 1995).

MLM enables the examination of relationships between data levels; for example, how a variable at Level-3 moderates a relationship between two Level-1 variables (cross-level interaction). It is possible to allow lower-level effects to vary at higher levels of the model. For example, a time predictor may, on average, be associated with an outcome (fixed effect) but this association may vary significantly between both days and participants (random effects).

MLM explicitly models the hierarchical data structure, leading to no lost information from aggregation or disaggregation procedures and therefore greater statistical power (Hox, 2010). MLM recognises the shared variance between assessments (dependency of observations) yielded from the same day and the same participant, accounts for the hierarchical data structure by assuming different relations for each assessment cluster (intra-class correlation coefficients), and is ideally suited for within-subjects, as well as between-subjects, analyses (Hruschka, Kohrt, & Worthman, 2005). Where EMA data were analysed in this thesis, MLM was chosen as the preferred statistical method.

As MLM is an extension of conventional linear regression, most assumptions hold for both methods. The main assumption of MLM is that the errors (residuals) at each level of the model follow a normal distribution.

However, Maas and Hox (2004) demonstrated that non-normality of residuals at higher levels did not affect fixed parameter estimates and lower-level variance, but may influence higher-level random effects if sample sizes have less than 10 groups. This was important to the present analyses, as SPSS cannot examine Level-2 or Level-3 residual distributions. Violations of residual variance constancy (heteroscedasticity) and uncorrelated residuals assumptions can be modelled in MLM via the specification of variance-covariance matrices.

2.2.6.1 Specification of variance-covariance structure

By definition, multilevel models have more error terms than normal regression models, and these error terms provide important information; namely, how much variability in the outcome is at Level-1 (between assessments), Level-2 (between-days) and Level-3 (between-individuals). One can combine Level-1 and Level-2 variability as a measure of within-individual outcome variance. It is also possible to examine the extent to which growth trajectories (change in outcome over time) vary across participants.

The flexibility in MLM afforded by multiple error terms permits a complex variance-covariance structure which improves the specification of the model and leads to better standard errors of the model parameters, more accurate confidence intervals, and a more valid statistical test (Dedrick et al., 2009). Efficacy of variance-covariance structures can be examined using fit indices such as the -2 Log Likelihood and Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics. Each chapter employing MLM in this thesis presents the model-build process and clearly states the variance-covariance structure used and how this was derived.

2.2.6.2 Centring

Predictor centring is vital to the interpretation of the intercept in MLM (Hox, 2010). With regard to the datasets in this thesis, assessment-level and day-level predictors were centred on the person-mean. In person-mean centring, the mean of the time-varying predictor is subtracted from the predictor value at each assessment point, leaving a predictor representing variation about an individual's mean (pure within-subjects variation). The interpretation of the parameter of the fixed effect thus becomes the increase in the outcome for every one-unit increase in the predictor above that which was usually experienced. In addition, the model intercept is interpreted as the level

of the outcome variable when the predictor is at its person-average (person-centred predictor = 0).

An exception was the *Time* predictor at Level-1 in the 3-level multilevel model. In this case, it was more appropriate to centre Time to a time of day making intuitive sense for intercept interpretation, so was usually centred to 1000h by subtracting 10 from the time variable (*CentredTime* = *Time* – 10). Empirical chapters clearly state where different time centring was performed.

2.2.6.3 Outliers

Potentially influential outliers were identified using the mean ±3*SD* rule, recommended for use in ambulatory psychoneuroendocrinology (Schlotz, 2011). However, scatter plots were always examined to "eyeball" data and interpret the likelihood of outlier influence. Where potential outliers were identified, sensitivity analyses were carried out to ensure results were not a consequence of influential outliers.

2.2.6.4 Missing assessments

MLM fits a slope and intercept to Level-1 units within Level-2, and the same for Level-2 slopes and intercepts within Level-3. This means, at minimum, two data units are required at each level and, with increasing number of units (assessments and days), parameter estimate precision also increases. MLM is relatively robust against missing data when using the Maximum Likelihood method, and is a further advantage over more conventional analyses where missing data usually results in case deletion and resultant loss of statistical power, which can itself introduce bias "unless the complete cases are a random subset of all cases in the sample" (Black, Harel, & Matthews, 2011, p.345).

There are assumptions to be met in MLM regarding missing assessments (missingness). Missingness becomes "ignorable" (MLM can be undertaken without modelling the non-response) when data are missing at random (Black et al., 2011; Graham, 2009). There are three theoretically-derived descriptions of missingness: (1) missing completely at random (MCAR); (2) missing at random (MAR); and (3) missing not at random (MNAR) (Graham, 2009). MCAR refers to missing data that is unrelated to observed or unobserved (missing) variables, whereas MAR is missing data accounted for by observed but not

unobserved variables. MNAR reflects missing data that depend on unobserved data. The missingness assumption is that missing data is MCAR or MAR.

Statistically testing whether missing data are MAR or MNAR is not possible because unobserved data is, by definition, unavailable; however, it is possible to test the more conservative MCAR assumption by examining whether plausible observed variables predict missing events. If so, then the MCAR assumption is rejected and predictors of missingness must be included in the model (Black et al., 2011). If not, there is no reason to assume data are not MCAR.

The best means of dealing with missing data is to avoid or reduce its occurrence (Black et al., 2011) and several strategies promoting compliance contributed to minimising missing data (see section 2.2.2.2.4). However, some missingness in EMA studies is inevitable due to being undertaken in everyday life (Black et al., 2011; Stone & Shiffman, 2002). Some arbitrary "rules of thumb" have been offered to ensure sampling representativeness, such as Stone and Shiffman (2002) who recommend a minimum 80% response rate for inclusion of a participant's data. This was not based upon any empirical evidence, but was rather a guideline issued by experts within the field of EMA research. However, this suggestion was not proposed for cortisol data where it often depends on the cortisol markers being computed as to how conservative compliance rates must be. As no empirically-resolved strategies for minimal response rates have currently been offered, and MLM appears particularly robust to missing values so long as they are MCAR or MAR, the strategy presented below was devised.

2.2.6.4.1 Missingness and non-compliance in CAR data

To reliably quantify the CAR, at least three samples are required to compute an area under the curve (Clow et al., 2004; J. C. Pruessner et al., 2003), meaning any missing morning sample (S1-S3) in the present study led to the exclusion of that day's CAR. Given that a Level-2 unit containing only one Level-1 unit would make the variance observed at each level indistinguishable (Nezlek, 2011), at least two valid CARs from the 4 assessment days were required or the participant's data was excluded.

Several studies have addressed the issue of "tolerable" delays in S1 samples post-awakening for CAR studies (Dockray, Bhattacharyya, Molloy, &

Steptoe, 2008; Okun et al., 2010; N. Smyth, Clow, Thorn, Hucklebridge, & Evans, 2013). Both earlier studies (Dockray et al., 2008; Okun et al., 2010) have argued an S1 sample delay of up to 15 minutes is tolerable; however, N. Smyth et al. (2013) demonstrated that cortisol levels were significantly elevated after just 10 minutes. No study has yet empirically examined tolerable delays in later assessments (S2 and S3), but Kudielka et al. (2003) previously set a ±7min criterion. Valid CARs were therefore defined as where S1 was provided within 10 minutes of awakening; S2 within 37 minutes of awakening, and S3 within 52 minutes of awakening.

Overall, 881 S1-S3 samples were provided out of a possible 912 (96.6%), with similar rates in the RRMS group (96.5%) and control group (96.7%). Timestamps indicated six S1 samples were provided more than 10 minutes after awakening, 13 samples more than 37 minutes after awakening, and 22 S3 samples more than 52 minutes after awakening. All of these delayed samples were removed. Valid CAR computations, requiring all three samples, were possible on 258 out of 304 assessment days (84.8%). In the RRMS group, valid CARs were provided on all 4 days in 63.2% of participants, at least 3 days in 89.5%, and at least 2 days in 97.4%. One RRMS participant was excluded from the CAR dataset as they provided no compliant CARs. In the control group, all 4 days were valid in 57.9% of participants, at least 3 days in 73.6%, and at least 2 days in 100%.

2.2.6.4.2 Missingness and non-compliance in DCS and momentary self-report

Growth models had the following missingness criteria: (1) completion of at least three Level-1 assessments out of six possible (A1–A6); and (2) at least 3 qualifying days to obtain reliable parameter estimates at Level-3. A DCS estimated upon only two cortisol samples may be unreliable due to effectively reflecting a straight line between two assessments, and biased due to the pulsatile fashion in which cortisol is secreted (E. A. Young et al., 2004).

Overall, 1661 out of a possible 1824 A1-A6 events were completed, representing an excellent overall compliance rate (91.1%) similar in both the RRMS (90.9%) and control (91.2%) groups. There was no statistically significant effect of event number on missingness in a binary logistic regression model ($\chi^2 = 5.88$, df = 5, $p = .32$), meaning that time of day did not influence

missingness. Omnibus chi square tests showed no evidence that individuals with higher levels of stress, negative mood, or fatigue (using momentary assessment person-means) were more likely to miss events (all *p*s > .25). To conclude, there was no evidence to suggest violation of the MCAR assumption.

2.2.6.5 Effect sizes

In MLM, the fixed effect parameter is a measure of effect size: the parameter equals the change in outcome for every one-unit increase in the predictor. However, to facilitate comparisons with previous research, Cohen's *d* (J. Cohen, 1988) effect sizes (formula and rationale in Appendix A) were computed using aggregated data based on independent samples t-tests.

2.3 Ethical considerations and funding

2.3.1 Participant burden and consultation with clinical group

Hufford (2008) described six aspects to consider when evaluating participant burden within an EMA design: (1) sampling frequency; (2) length of each assessment event; (3) cognitive effort necessary for completion; (4) overall assessment duration; (5) reporting platform usability; and (6) reporting platform reliability (p.65). Each of these points should be considered against the need, both empirically and ethically, to measure phenomena appropriately. However, participant burden was a major consideration within a clinical population often presenting with cognitive and motor difficulties (Rao, Leo, Bernardin, & Unverzagt, 1991).

To scrutinise the suitability of the EMA protocol for people with RRMS, two pwRRMS meeting study eligibility criteria were individually consulted on the intended protocol, equipment, and supportive documents before submission to ethics committees and governance bodies. Each of the six points regarding participant burden were addressed in the consultation, and each individual was talked through the demonstration version of the handheld programme. Both individuals felt participant burden was not unreasonable given the study aims, particularly given it would be completed after 4 days. The handheld was considered intuitive, and the time taken to complete the longest events (A1-A6; ~2 minutes) was acceptable. Both individuals agreed the EMA protocol was feasible in a remission-phase RRMS population. The

main points emerging from the consultations requiring assurances and/or amendments to the protocol are described in Table 5.

Table 5 Comments requiring Remedial Action made in Consultations with People with RRMS

Theme	Comment	Action taken
Privacy	Some pwMS do not tell employers, friends, or family of their diagnosis. It is essential that participation does not reveal/publicise an MS diagnosis.	The design was identical for both groups. The protocol contained no features identifying MS.
Practicality	Although the handheld font is readable, it would be prudent to make the font as large as possible.	Font was enlarged as much as possible while retaining all item wording on one screen.
	Documentation should be available in large- and normal-font size, with participants able to choose their preferred medium.	Documentation was provided as suggested.

2.3.2 Ethical approvals

Following the consultation, ethical approval was sought. Full ethical approval was obtained from the University of Southampton Psychology Ethics Committee (*Study Number 589*), University of Southampton Research Governance Office (*Study Number 8031*), the NHS National Research Ethics Service Committee South Central (*Reference Number 11/SC/0333*), and the Research and Development departments at University Hospital Southampton NHS Foundation Trust and Guys and St Thomas' NHS Foundation Trust to permit access to MS clinics for recruitment purposes.

Before deciding whether or not to take part, all potential participants received a PIS and a 48-hour period elapsed before follow-up contact was made. All links between data collected and personal information were broken using a random code string known only to the researcher.

All participants provided written informed consent, and received £40 reimbursement for their time, effort, and expenses incurred taking part in the study. This level of reimbursement was considered justifiable as the protocol extended over several days and required two visits to the University of Southampton or Kings College London. Reimbursement was fixed at £40 for total transparency to each participant, and ease of administration.

2.3.3 Research funding

Participant reimbursement, equipment, postal, and cortisol analysis costs were funded by the Psychology Unit at the University of Southampton. The researcher was funded by a PhD Studentship from the Economic and Social Research Council (Ref: ES/1026266/1).

2.4 Chapter summary

EMA is a contemporary methodological strategy enabling the observation of phenomena in the real-world and in real-time, while permitting within-subjects examinations of contextual effects. The chapter outlined the rationale for using an EMA strategy in this thesis, and presented the recruitment procedure, participant characteristics, design, and procedure for this doctoral research. In the empirical chapters to follow, relevant features of the methodology are presented, in brief, for convenience, and make reference to sections in Chapter 2 where appropriate. Chapter 2 also presented the advantages to utilising MLM for EMA data; unique multilevel models for testing specific hypotheses are specified in respective empirical chapters.

Chapter 3: Phenomenology of MS fatigue in everyday life

Despite the high prevalence rates for fatigue (60-80%) in MS (Hadjimichael et al., 2008; Lerdal et al., 2003; Minden et al., 2006) little is known about MS fatigue variability within days. This chapter presents a prospective examination of diurnal fatigue patterns and the within-subjects contextual effects of daily life stressors and mood using EMA in pwRRMS and healthy controls. Finally, an analysis is presented investigating the accuracy of end-of-day (EOD) recalled fatigue and the presence of recall bias in relation to momentary fatigue (MF) measures.

3.1 Introduction

3.1.1 Diurnal fatigue patterns in MS

MS fatigue has been described qualitatively as typically most severe in the late afternoon or early evening (Freal et al., 1984; Mills & Young, 2008). A diurnal fatigue pattern refers to the typical change in fatigue severity experienced by individuals over the course of a day. EMA offers a means of prospectively observing diurnal fatigue patterns, but current research on within-day fatigue variability in MS has primarily been conducted in specialised settings (for example, outpatient or inpatient clinics) where ecological validity is limited and comparison groups are missing.

Only one study has examined momentary reports of fatigue in MS in everyday life (E. Kim et al., 2010). In this study, actiwatch wrist devices (Mini Mitter, Respironics*, Bend, OR*) prompted MF ratings 4 times per day at fixed intervals (0900h, 1300h, 1700h, 2100h) for 21 days in 49 pwMS (39 RRMS, 7 SPMS, 3 PPMS). MF was rated by VAS from 0 (no fatigue) to 10 (worst possible fatigue) via the watch-face. MF scores were aggregated across days and across participants for each time-point, and represented a fairly linear increase in fatigue throughout the day (0900h, mean fatigue M_f = 3.4; 1300h, M_f = 4.1; 1700h, M_f = 4.5; 2100h, M_f = 5.0) with statistically significant differences between all time-points ($p < .01$). The effect of time (the diurnal pattern) was not explicitly examined.

The implied linear trajectory supported previous qualitative assertions on the diurnal fatigue pattern in MS (Freal et al., 1984; Mills & Young, 2008). However, without a healthy control group, it was not possible to discern whether diurnal fatigue patterns in MS are any different to that experienced by healthy people. Compliance in the E. Kim et al. (2010) study was relatively low, with 36% of Actiwatch reports missing, which likely biased the findings as ratings were likely missed more frequently when busy or tired.

Four other studies have used a design permitting examination of diurnal fatigue patterns in MS; but all were carried out in clinical settings or as part of an RCT (Claros-Salinas et al., 2010; Feys et al., 2012; Morris, Cantwell, Vowels, & Dodd, 2002; Schwid et al., 2003). The most measurement-intensive study was within an RCT for cooling therapy in 84 pwMS (78 RRMS, 6 SPMS/PPMS) with MF measured hourly using the Rochester Fatigue Diary (Schwid et al., 2002) as a secondary outcome over 24h periods on 8 days dispersed over 1 month (Schwid et al., 2003). The Rochester Fatigue Diary is a vertical-line VAS ranging from 0 (maximum fatigue/no energy) to 100 (no fatigue/maximum energy). The effect of time on fatigue was not explicitly examined, but descriptive information about the average diurnal fatigue pattern across participants could be derived from the presented figure (p.1959). While typical fatigue increased from late morning until the evening, this increase decelerated in the late afternoon before accelerating again from around 1900h-2100h. “Spikes” in fatigue were evident early in the morning and very late at night, but likely reflected sleepiness rather than fatigue. No information was presented about missing assessments.

The remaining studies were carried out in clinical settings. The Rochester Fatigue Diary was used in a study of 102 MS outpatients (53 RRMS, 31 SPMS, 18 PPMS) measuring MF over 1 day in the morning (0900h–1000h), noon (1200h–1300h), and afternoon (1500h–1600h) (Feys et al., 2012). Similar to previous studies, energy levels decreased (fatigue increased) over time (mean energy in morning: 73.5; noon: 52.3; afternoon: 51.7). However, there was an apparent deceleration in the reduction in energy later in the day; average energy decreased by 21.2 units from morning to noon, compared to 0.6 units from noon to afternoon. Furthermore, changes in fatigue scores over time did not differ between those with mild (1.5 < EDSS < 4.0) and moderate (4.5 < EDSS < 6.5) ambulatory dysfunction.

Claros-Salinas et al. (2010) measured MF ("How mentally fit do you feel at the moment?" on a 1-10 VAS) over 2 days at morning (0745h-0845h), noon (1130h-1230h), and afternoon (1545h-1645h) in 20 MS inpatients (type unclear), 22 stroke inpatients, and 76 healthy controls. However, only MF ratings pooled across all groups were reported, and increased with time: morning, M_f = 4.2; noon, M_f = 4.6; afternoon, M_f = 5.1. MS-specific data were not reported, but a relatively linear fatigue trajectory over time was evident in the figure presented (p.77). No statistically significant group by time interactions were evident, meaning diurnal fatigue patterns were similar across groups. However, group comparisons should be interpreted with caution due to different measurement contexts; healthy controls provided MF ratings while undertaking usual daily activities whereas the clinical groups were all inpatients.

The remaining clinic-based study measured MF ratings and gait at 1000h and 1500h in 14 pwMS (type unclear) on a single day (Morris et al., 2002). Ratings were provided using a single-item VAS ranging from 0 (no perceived fatigue) to 10 (worst fatigue ever experienced). MF increased from 1000h (M_f = 2.50 ±2.43) to 1500h (M_f = 3.75 ±2.75) with a statistically significant difference.

It would seem MS fatigue typically increases with time within-days, with most intensive fatigue severity occurring in the late afternoon, supporting earlier retrospective and qualitative findings (Freal et al., 1984; Mills & Young, 2008). However, conclusions must be drawn tentatively due to only 2-3 momentary assessments obtained per day in most studies. There are several general limitations to note. Firstly, fixed-occasion designs have been used in all studies, limiting generalisability as they do not provide representative samples of daily experience. Statistical analysis explicitly examining the relationship between fatigue and time is lacking in studies, with fatigue often only a secondary outcome. Heterogeneous MS samples have also been recruited and, despite fatigue levels being shown to be significantly different between MS types (Kroencke et al., 2000), subgroup analysis has not been performed. Most studies also lack ecological validity, particularly those recruiting inpatient samples where nocturnal awakenings are typically increased (J. S. Young, Bourgeois, Hilty, & Hardin, 2008).

3.1.2 Contextual effects of stress and mood

Investigating factors influencing diurnal fatigue patterns in clinical populations may lead to useful insights into its aetiology (Curran et al., 2004). Simmons et al. (2004) reported 78% of 2529 pwMS believed stress worsened symptoms, including fatigue. In another survey, 90% of 120pwMS responded "yes" to the question "Does experiencing a stressful situation increase your fatigue?" and 87% replied "yes" to "Does an increase in your daily activities increase your fatigue?" (Mollaoğlu & Üstün, 2009).

Associations have also been reported based on retrospective questionnaires. In a study of 53 pwMS (RRMS/SPMS), self-reported stress (PSS; S. Cohen et al., 1983) was associated with fatigue on both the Multidimensional Fatigue Inventory (Smets et al., 1995) (r = .35) and the FSS (r = .34) (Trojan et al., 2007). R. F. Brown et al. (2009) measured fatigue impact (Fatigue Impact Scale; Fisk, Ritvo, et al., 1994) and stress (Life Events and Difficulties Schedule; G. W. Brown & Harris, 1989) every 3 months for 2 years in 101 pwMS (RRMS/SPMS), and demonstrated an accumulation of emotional threat was associated with later fatigue (γ = 0.98, p < .001). However, the number of participants completing the 2-year protocol was very small (25%), and drop out may have biased results.

These findings suggest stressful experience is a potential moderator of fatigue severity in MS, but only permit inferences that consider summarised and retrospectively appraised levels of perceived stress and fatigue. The literature lacks a within-subjects investigation examining whether fluctuations in everyday stressors elicit effects on concurrent fatigue severity in MS. Several qualitative explorations have touched on the immediacy of any effect of stress on fatigue for pwMS; for example, Stuifbergen and Rogers (1997) quoted a participant stating: "Stress can really make you feel tired… on hard emotional tasks, at the end of them, I feel really drained" (p.6). In a qualitative study of barriers to employment a derived theme described stress impacting upon symptoms; for example, "When I get stressed I can't talk very well… it makes my MS worse immediately" (Johnson et al., 2004 , p.205).

Similarly, cross-sectional studies have found associations of depression and anxiety with MS fatigue (Bol et al., 2009) but have not yet been able to extrapolate whether mood is associated with fatigue on a moment-by-moment

basis within-subjects. Research is hindered by the fact fatigue is itself a common symptom of depression (DSM-IV, American Psychiatric Association, 1994). A common solution has been to omit items from depression scales that refer to fatigue symptoms. In a study of 739 pwMS (Chwastiak et al., 2005), 76% of individuals with disabling fatigue (⩾15 on MFIS; Ritvo et al., 1997) also had clinically-significant depressive symptoms, based on a score ≥16 on the Centre for Epidemiological Studies Depression Scale (Radloff, 1977) with fatigue items removed; whereas 31% without disabling fatigue reported clinically-significant depressive symptoms. In a multivariate model of fatigue containing neurological disability (EDSS), secondary progressive disease course, and depression score as predictors, depression accounted for 23% of variance.

A study with 25 breast cancer survivors (6–26 months post-treatment) and 25 healthy controls represents the only study of the contextual effects of daily life factors (in this case, mood) on fatigue in a clinical population (Curran et al., 2004). Momentary assessments of fatigue, mood, and pain were obtained over 5 consecutive days in everyday life at waking, 1000h, 1400h, and 2100h. However, the analysis did not represent a within-subjects analysis as fatigue and mood were aggregated for each time-point across the 5 days and a between subjects analysis was performed for each time point. In the cancer survivors group, there was a small negative momentary association between positive mood and fatigue (all *r*s between -.19 and -.29, *p*s < .05) and a slightly larger positive associated between negative mood and fatigue (all *r*s between .27 and .54, *p*s < .05). Causality was unclear. There has been no empirical examination of within-person within-day associations of everyday life psychosocial experience with MS fatigue.

3.1.3 Accuracy of recalled fatigue

The vast majority of fatigue measures are traditional paper-pen questionnaires requiring the recollection of fatigue experience over a given period of time. In reality, once experience has ceased it can no longer be directly evaluated and individuals reconstruct experience by evaluating a series of past experiences to formulate a meaningful response. Retrospective self-reports often inform clinical diagnostic and therapeutic decisions, meaning their accuracy and true representativeness is central to appropriate health

management. However, affective experiences such as pain or negative mood are often far from straightforward to quantify over a prolonged period given relatively frequent fluctuations in experience intensity. The accuracy of recalled fatigue reports in MS has not been examined.

3.1.3.1 Discrepancy between reporting periods

Typically, studies comparing retrospective measures of somatic symptoms to aggregated momentary ratings describe overestimation in recall. For example, Stone et al. (2004) requested momentary pain ratings and weekly-recalled pain ratings over 2 weeks in a group of 68 individuals experiencing chronic pain. Momentary pain was measured by "How much pain are you in right now?" (0-100 VAS) presented on a small handheld device, alongside weekly recalled pain ratings measured by indicating usual pain on a 0-100 VAS during a visit to the research office. Statistically significant over-estimations were evident in recalled pain compared to aggregated momentary pain ratings in week 1 (57.7 ±22.3 vs. 44.4 ±22.1) and week 2 (59.7 ±21.1 vs. 44.3 ±23.2). After excluding periods of no pain (pain = 0) from momentary aggregates, discrepancy was reduced by 60%, indicating participants tended to consider only periods where pain was present in recall provision. The framing of the weekly recall question ("usual pain") may have steered responders towards considering only moments where pain was present.

Although paper rather than electronic VAS were used in weekly pain reports, a previous randomised study in healthy participants revealed remarkably similar ratings between a 10cm paper VAS and a 5cm electronic VAS via handheld device (r = .91) (Jamison et al., 2002). It was possible a researcher was present (not confirmed in the report) during the provision of weekly ratings in the study by Stone et al. (2004) which may have introduced social desirability bias not present with device responses.

Recall overestimation has also been reported elsewhere. Peters et al. (2000) found a statistically significant difference in pain ratings on the multidimensional pain inventory (Kerns, Turk, & Rudy, 1985) between recalled (M = 4.0, representing "horrible") and aggregated momentary scores (M = 2.8, representing somewhere between "discomforting" and "distressing") in 80 individuals with unexplained pain. Van Den Brink, Bandell-Hoekstra, and Abu-Saad (2001) found headache intensity was overestimated in retrospective

questionnaires (median intensity: 65 on VAS) compared to 4-week daily diaries (median intensity: 37 on VAS) in a paediatric population experiencing frequent headaches (n = 181).

Houtveen and Oei (2007) examined recall discrepancy in medically unexplained symptoms (MUS) in undergraduate females with high (n = 18) and low (n = 19) levels of MUS, defined as being above the 75th percentile (high) or below the 25th percentile (low) on the somatic subscale of the Symptom Checklist (Derogatis, 1977). Handheld devices were used to make stratified assessments of symptoms (mean 6.4 alarms per day) over 7 days, with additional day-recall and week-recall reports. Upon each alarm, participants rated 12 symptoms on how they felt at that moment on a 7-point Likert Scale ranging from "not at all" to "extremely". At the end of each day, participants completed the same 12 items framed to summarise daily experience, and at the end of 7 days the items were framed to summarise the week. Recall bias was operationalized as (1) week-aggregated day-recall minus week-aggregated momentary-recall (day recall bias); and (2) week-recall minus week-aggregated day-recall (week recall bias).

Both high-MUS and low-MUS groups had higher week-recalled and day-recalled symptoms compared to aggregated momentary symptoms (mean and *SD*s not provided), indicating the existence of recall bias even over relatively short periods. Similar discrepancies were present for pain-related and nonspecific complaints. Daily recall bias (overestimation) was associated with momentary *SD* (greater variability) in individuals with low MUS (r = .50, p = .03) but not high MUS (r = .24, p = .34). It was not possible to infer causality, but more variation in symptom intensity experienced by those not well-accustomed to such experiences may lead to inaccuracies in retrospective reporting.

The importance of within-person symptom variability to recall discrepancy was further demonstrated by Sohl and Friedberg (2008) who asked 53 people with chronic fatigue syndrome (CFS) to provide MF ratings 6 times per day for 21 days using an electronic diary. Participants responded to the item "Fatigue Now" on a Likert scale from 0 ("none") to 10 ("highest"). At the end of each week, participants were telephoned to obtain weekly fatigue ratings using the same scale but verbally: a potential source of bias as participants may be more or less likely to over-report symptoms verbally over

the telephone versus via an electronic device. Recall discrepancy was calculated by subtracting week-aggregated momentary ratings from weekly recall ratings, and fatigue variability was represented by the *SD* of all momentary ratings. Results revealed a small but not quite statistically significant over-reporting of fatigue in recall ratings (M_f = 5.19 ±1.63) compared to momentary ratings (M_f = 4.93 ±1.73, d = 0.15, p = .07).

Unlike in the study by Houtveen and Oei (2007), where symptom variability appeared only relevant to recall discrepancy in those who perhaps did not experience symptoms frequently, Sohl and Friedberg (2008) found an effect of fatigue variability on recall discrepancy in a CFS population; recall discrepancies were associated with fatigue variability in week 1 (r = .43, p < .01), had a tendency towards a statistically significant association in week 2 (r = .27, p = .07), but no relationship in week 3 (r = .07, p =.65). These results indicated a possible practice effect. Recall discrepancy was also negatively associated with mean fatigue intensity (r = -.53, p <.001), so those with higher fatigue levels during each day provided more accurate recall, and with depression scores on the Beck Depression Inventory (BDI-II; Beck et al., 1996) (r = -.30, p < .05), such that higher levels of depression were associated with lower recall discrepancies. The inverse relationship with depression may indicate greater somatic awareness in individuals with depression.

The studies reviewed all suggest recalled symptom ratings are not merely perceived averages of intensity over a given time period, but are influenced by other factors or heuristics, such as the variability of the pain, and potentially the salience of the pain. This latter point is often referred to as a peak heuristic, and has been the focus of considerable research examining recall bias, along with the end (or recency) heuristic and is discussed in greater detail in the next section. Tentatively, one may expect the practice effect observed by Sohl and Friedberg (2008) to also manifest as more accurate retrospective symptom reporting in clinical groups compared to nonclinical groups due to greater familiarity with symptoms (they may attend to it more readily) and greater experience in the provision of such reports as part of routine interactions with health professionals. However, there has yet to be a case-control study examining discrepancies between momentary and retrospective reports of somatic symptoms in which this hypothesis could be tested.

There is an argument that recall overestimations are often negligible and of little clinical significance. Indeed 18-20 units on a 0-100 VAS has been described as a minimal clinically important difference for pain ratings (Grilo, Treves, Preux, Vergne-Salle, & Bertin, 2007; Hägg, Fritzell, & Nordwall, 2003); a size of overestimation, for example, that the Stone et al. (2004) study failed to observe despite statistical significance. The discrepancy found by Sohl and Friedberg (2008) was also unlikely to be clinically significant, with only a 2.6% difference between average weekly and momentary ratings.

3.1.3.2 Peak and end biases in recall

Recall inaccuracies are not thought to be due to random noise, but instead to systematic biases such as the disproportionate importance given to most intense or salient (peak) and most recent (end) experience in formulating responses (Fredrickson, 2000). An early seminal study by Kahneman, Fredrickson, Schreiber, and Redelmeier (1993) informed 32 male students they would be undertaking three trials involving hand submergence in cold water. Trial 1 was 60 seconds in 14°C cold water; trial 2 was 60 seconds in 14°C cold water followed by gradually raising the temperature to 15°C for a further 30 seconds; and trial 3 was fictitious. Trial ordering (1 and 2) was balanced across participants and participants immersed both hands in room-temperature water for 2 minutes before each trial. Discomfort was measured every 5 seconds using a "discomfort meter": a row of 15 small LED lights indicating increasing discomfort (scored 0-14). This measure has uncertain psychometric properties. After both trials, participants considered the discomfort of each trial and indicated which they would rather repeat for "trial 3".

Momentary discomfort ratings were similar for the first 60 seconds of both trials (trial 1, M = 8.44; trial 2, M = 8.34), and discomfort levels dropped by a mean of 2.65 for the final 30 seconds of trial 2 (still indicating discomfort). Most participants reported trial 1 elicited the most discomfort, was most difficult to cope with, and 69% of participants preferred to repeat the longer trial 2. The authors argued this decision represented "duration neglect" whereby pain duration was ignored in experience reconstruction in favour of both peak and end heuristics. Average momentary discomfort ratings were above 4 throughout both trials, suggesting numbing of the hand did not occur, although this cannot be completely discounted.

EMA has permitted the investigation of peak and end effects in everyday life. Stone et al. (2000) collected momentary pain reports over 7 days in 35 people with rheumatoid arthritis on a 7-point Likert Scale from "not at all" to "extremely". Momentary reports were prompted by quasi-random wristwatch beeps seven times per day between 0800h and 2100h. On the eighth day, while in the rheumatologist's office awaiting their routine appointment, participants recorded a summarised pain rating for the week on a 10cm VAS anchored at "no pain" and "pain as bad as it could be". Peak and end biases were combined by averaging peak (highest momentary score) with end1 (average final day pain) and end2 (highest final day pain).The authors did not report overestimation or underestimation of weekly pain reports, presumably as it is problematic to do so with different response scales completed in different contexts (Broderick et al., 2008).

Weekly pain reports were associated with peak-end1 (r = .78, p < .05), peak-end2 (r = .80, p < .05) and average momentary pain reports (r = .72, p < .05). Average pain was then entered into hierarchical multiple regression at step 1 followed by one of the peak-end combinations. Peak-end1 accounted for an additional 8.2% of the variance (p < .05), whereas peak-end2 accounted for an additional 11.5% of the variance (p < .01). When the stepwise order was reversed, average pain accounted for only 0.1% of variance (p = ns) with peak-end1 at step 1, and 0.6% of variance (p = ns) with peak-end2.

EOD ratings tend to provide more accurate representations of pain experience than retrospective 7-day reports (Broderick, Schwartz, Schneider, & Stone, 2009), suggesting heuristic biases may be more evident in longer recall periods. However peak-end biases have been demonstrated in EOD recall of pain, both between- and within-subjects (Schneider, Stone, Schwartz, & Broderick, 2011). Over 28 consecutive days, 7 quasi-random prompts requested momentary pain ("Before the prompt: How intense was your bodily pain?") and fatigue ("Before the prompt: How fatigued (weary, tired) did you feel?") ratings on an electronic device in 97 rheumatology patients. EOD ratings were provided when going to bed by manually engaging the handheld: "What was your usual level of your pain/fatigue today?" All responses were via 0-100 VAS from "not at all" to "extremely". Peak fatigue/pain and end fatigue/pain were operationalized as deviations from daily mean fatigue/pain.

In between-subjects analyses, mean pain ratings explained 84% of EOD pain variance, with peak and end levels (separate variables) explaining an additional 2% ($p < .01$). Mean fatigue ratings explained 76% of EOD fatigue variance, with peak and end levels not making a statistically significant contribution to the model. In within-subjects analyses, mean pain ratings explained 63% of EOD pain variance, with peak and end levels explaining an additional 3% ($p < .01$). Mean fatigue levels explained 52% of the variance in EOD fatigue, but peak and end levels were not a significant contributor to the model. The study provided evidence that the peak and end heuristics have a biasing influence over EOD pain reports, but this effect is minimal (2-3%, between- and within-subjects) and not apparent at all for fatigue reports.

Similar results were found for EOD pain ratings in 1062 post-operative inpatients (Jensen, Mardekian, Lakshminarayanan, & Boye, 2008). Verbal rating scales for pain were completed on a 4-point Likert Scale ranging from 0 ("none") to 3 ("severe") 5 times per day over 10 days, with EOD reports on a 0-10 VAS from 0 ("no pain") to 10 ("pain as bad as you can imagine"). On each day, average momentary pain accounted for 44-55% of EOD pain variance, and peak and end pain scores accounted for small, yet statistically significant, additional variance (2-4%, all ps $< .001$).

The two studies examining peak and end biases in EOD reports indicate only a small biasing compared to weekly reports, where peak and end bias accounted for 8%-11% of variance in addition to that accounted for by average ratings (Stone et al., 2000). Schneider et al. (2011) failed to observe any peak and end effects in EOD fatigue reports, which was surprising given the relationship between mean fatigue ratings and EOD fatigue was relatively small (explaining only 63% of between-subjects EOD variance, and 52% within-subjects, compared to 84% and 76% of respective EOD pain variance) which may suggest heuristics "potentially play a larger role in fatigue recall than in pain recall" (Schneider et al., 2011, p.233). Schneider et al. (2011) speculate that recalled fatigue may be more complex than pain, with the impact of fatigue on the individual's ability to complete their daily tasks potentially influencing EOD reports over and above peak or end heuristics.

The majority of studies examining peak and end biases in recalled symptoms have been in rheumatology populations, and it is unclear whether findings generalise to other populations with prominent somatic symptoms,

such as MS. Recall biases have not always been evident in the few studies in other clinical populations (for example in neck pain patients' weekly pain assessments; Bolton, Humphreys, & van Hedel, 2010). It is not clear whether clinical populations differ from healthy individuals in their use of peak and end memory heuristics, although there is evidence depressed patients use them more readily than healthy individuals when recalling daily negative affect (Ben-Zeev, Young, & Madsen, 2009). One may expect differences between pwMS and healthy individuals due to practice effects borne from patients' familiarity with the symptom and relatively frequent summarising of somatic symptoms to health practitioners at routine clinics; however, this hypothesis has yet to be tested.

3.1.4 Aims

The overall aim of this chapter was to investigate how fatigue is experienced and reported in an RRMS group compared to a healthy control group. The diurnal fatigue pattern was examined in a case-control design, with the following hypotheses tested: (1) average levels of fatigue in everyday life are higher in pwRRMS than healthy controls; (2) average fatigue slopes (diurnal fatigue patterns) have steeper increases in pwRRMS than controls; (3) across groups, increases in prior stressors are associated with increased MF; (4) across groups, increases in distress are associated with concurrent increases in MF; (5) across groups, increases in positive mood are associated with concurrent decreases in MF.

Research has suggested recall bias may be present in somatic symptom reporting, but it is unclear whether these findings generalise to fatigue recall in RRMS. Discrepancy between EOD fatigue ratings and mean MF assessments in RRMS was examined, as well as peak and end bias. Several hypotheses were tested: (1) in both groups, fatigue is overestimated by EOD fatigue ratings compared to aggregated MF scores; (2) discrepancy is smaller in pwRRMS than healthy controls; (3) across groups, symptom variability is positively associated with recall discrepancy; (4) across groups, discrepancies are smaller with each assessment day that passes (practice effect); (5) daily peak and end fatigue ratings account for EOD fatigue variance beyond that of average MF ratings; and (6) pwRRMS employ the peak and end heuristic less than controls.

3.2 Method

Rationale and general methods were presented in section 2.2 of Chapter 2. For convenience, relevant aspects of the method are presented in brief. Salivary cortisol assessment was not relevant to this chapter. Section 3.2.4 below provides a detailed overview of the statistical procedures carried out to test hypotheses.

3.2.1 Participants

The study recruited 38 individuals (31 female, 7 male) to an RRMS group and 38 individuals (31 female, 7 male) to a healthy control group. Age ranged from 27–56 years (*M* = 41.89 ±7.53 years) in the RRMS group and 28–54 years (*M* = 40.34 ±8.16 years) in the control group. Eligibility criteria and recruitment procedures were detailed in section 2.2.1 of Chapter 2.

Full descriptive statistics of the study sample was provided in Table 2 in Chapter 2 (p.51). Most of the RRMS group (52.6%) were in full-time (>31 hours/week) paid employment, 26.3% part-time, and 21.1% unemployed. In the control group, 63.2% worked full-time, 21.1% part-time, and 13.2% unemployed (one individual was in paid employment but did not indicate usual weekly hours).

3.2.2 Design

Participants attended an introductory session where they received full training in the use of the handheld (Hewlett Packard iPAQ 111 Classic Handheld, *Bracknell, UK*) prompting the EMA protocol, and completed baseline questionnaires. EMA then took place over 4 consecutive weekdays while participants undertook usual daily routines. The six quasi-random events prompted between 1000h and 2000h (A1–A6) and the EOD event at 2100h were of relevance to this chapter. The event-related morning design (events S1–S3) did not provide data for this chapter, except for perceived sleep quality measured at S1 (awakening).

3.2.3 Measures

3.2.3.1 Baseline measures

Of relevance here were the Expanded Disability Status Scale (EDSS; Bowen et al., 2001); the Fatigue Scale (FS; Chalder et al., 1993); the short-form Modified Fatigue Impact Scale (MFIS-5; Ritvo et al., 1997); and the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D; Zigmond & Snaith, 1983). Socio-demographic and other personal variables were also assessed for examination as potential covariates. Details of baseline questionnaires, and their validities and justifications were discussed in section 2.2.3.1 of Chapter 2.

3.2.3.2 EMA measures

All responses to EMA self-report items were provided by dragging an indicator to a position on a 0–10 VAS displayed on the handheld screen. Sleep quality was assessed upon awakening by "How would you rate the quality of your sleep last night?" with responses anchored at "Very Bad" (0) to "Very Good" (10). The below were presented at each quasi-random event (A1–A6) on all 4 days.

3.2.3.2.1 Momentary fatigue

MF was measured by the single-item question "How much fatigue (weariness, tiredness, problems thinking clearly) do you feel right now?" anchored at "None at all" (0) to "Extreme fatigue" (10). This item was developed from the Brief Fatigue Inventory (Mendoza et al., 1999) with the addition of the words "problems thinking clearly" to parentheses to reflect the mental aspect of MS fatigue definition (Multiple Sclerosis Council for Clinical Practice Guidelines, 1998).

3.2.3.2.2 Momentary stressors

Stressors were measured by eight items preceded by the phrase "Since the last event…" encapsulating all stressors experienced during the day. Items included "I did a lot of work" and responses were anchored at "Not at all" (0) and "Very much so" (10). The eight items were reduced to two factors by an exploratory factor analysis (see section 2.2.3.2.3.2 of Chapter 2): *Stressor Challenge* (four items) and *Stressor Hindrance* (four items). Challenge was

conceptualised as effortful demands that may lead to goal achievement and stimulation, and Hindrance was conceptualised as demands that are potential barriers to goal achievement. Each factor was scored as the mean of constituent items.

3.2.3.2.3 Momentary mood

Mood was measured by 14 items preceded by the phrase "At the moment I feel…" Items included "Upset" and "Determined" and responses were anchored at "Not at all" (0) and "Very much so" (10). The 14 items were reduced to two factors by exploratory factor analysis (see section 2.2.3.2.3.4 of Chapter 2): *Distress* (9 items) and *Positive Mood* (5 items). Each factor was scored as the mean of constituent items.

3.2.3.3 End-of-day fatigue measures

At 2100h, daily fatigue was retrospectively appraised using a single item reflecting the entire day: "How much fatigue (weariness, tiredness, problems thinking clearly) have you felt today?" Responses were anchored at "None at all" (0) and "Extreme fatigue" (10).

3.2.4 Statistical analysis

Six assessment events (A1–A6; Level-1) were nested within 4 days (Level-2) nested within 76 participants (Level-3). Group (control or RRMS) was indicated by a Level-3 dummy variable (0, 1). To account for the non-independence of observations and missing observations within the data, analysis was by multilevel modelling (MLM) using full information maximum likelihood estimation (S. Liu, Rovine, & Molenaar, 2012; Singer & Willett, 2003). Due to the variable-occasion EMA design, standard growth curve models included time parameters as fixed effects at Level-2 and Level-3 and these effects (slopes) could vary at both levels (random effects) to account for non-independence of observations within individuals (intra-class correlations). All specified models follow the notation reported by Snijders and Bosker (2012). Effect sizes were computed as Cohen's *d* using aggregated data (see Appendix A). Analyses were carried out using SPSS Version 20.0. Statistical significance was defined by the $\alpha = .05$ criterion.

3.2.4.1.1 Data cleaning

Participant compliance was checked using recommended criteria for everyday life studies with electronic platforms (McCabe, Mack, & Fleeson, 2011). Non-compliance was defined as event responses meeting any of the following criteria: (1) item-response too fast (< 0.5 seconds); (2) responses too similar (>90% identical responses to items within event); (3) event completion too slow (>mean time +3*SD* = 6 minutes, 46 seconds). Criterion 1 indicated insufficient comprehension of an item for an unbiased response (removed items = 0). Criterion 2 identified events containing an implausible level of identical responses to items (removed events = 7). Finally, criterion 3 eliminated substantial delays between momentary assessments threatening the validity of momentary associations (removed events = 18; see Appendix I).

3.2.4.2 Building the MLM for diurnal fatigue patterns

To facilitate interpretation of model intercept, time was centred at 1000h (*CentredTime = Time of Assessment (hours) – 10*) so the intercept represented mean MF at 1000h. To model change in MF, linear and quadratic growth models were explored for best-fit. Growth models were built and applied to the dataset in the following order: (1) unconditional null model; (2) random linear model; (3) fixed quadratic random linear model; (4) random quadratic model; (5) random quadratic model with quadratic effect permitted to vary at Level-2 only; (6) covariate growth model. Every model is specified in Appendix J and uses an *unstructured* variance-covariance matrix. Model parameter estimates are presented in Table 6.

Table 6 Multilevel Estimates for Baseline Polynomial Growth Model for Fatigue with covariates

		Parameter	Model 1 Unconditional	Model 2 Random Linear	Model 3 Fixed Quadratic Random Linear	Model 4 Random Quadratic	Model 5 Random (Level-2 only) Quadratic [a]	Model 6 Random (Level-2 only) Quadratic [a] Covariate
Fixed Effects								
Composite Model	Intercept (fatigue at 1000h)	γ_{000}	3.76***	2.63***	2.33***	2.36***	2.35***	2.24***
	TIME	γ_{100}		0.23***	0.40***	0.38***	0.38***	0.40***
	$TIME^2$	γ_{200}			-0.02**	-0.01~	-0.01~	-0.02*
	EXERTION	γ_{300}						0.69***
	SLEEP QUALITY	γ_{010}						-0.09*
Variance components								
Level-1	Within-day	σ^2_R	2.79***	1.83***	1.81***	1.53***	1.53***	1.52***
Level-2	Intercept	σ^2_0	0.78***	1.93***	1.95***	2.51***	3.02***	2.95***
	Linear term							
	Variance	σ^2_1		0.04***	0.04***	0.53**	0.67***	0.63***
	Covar with intercept	σ_{01}		-0.20***	-0.20***	-0.80*	-1.06***	-1.03***
	Quadratic term							
	Variance	σ^2_2				0.01***	0.007***	0.006***
	Covar with intercept	σ_{02}				0.07*	0.09**	0.09**
	Covar with linear term	σ_{12}				-0.05***	-0.07***	-0.06***
Level-3	Intercept	σ^2_{00}	5.43***	5.32***	5.32***	5.38***	4.98***	4.90***
	Linear term							
	Variance	σ^2_{10}		0.04***	0.04***	0.21~	0.04***	0.04***
	Covar with intercept	σ_{0010}		-0.09	-0.09	-0.25		
	Quadratic term							
	Variance	σ^2_{20}				0.001		
	Covar with intercept	σ_{0020}				0.008		
	Covar with linear term	σ_{1020}				-0.014		
Goodness of fit	-2 Log Likelihood		6742.74***	6407.83***	6400.67***	6356.24***	6364.06	6246.56***

[a] Variance-covariance at Level-3 has Diagonal structure (covariance of linear term with intercept not estimated due to a lack of statistical significance)

*** $p < .001$, ** $p < .01$, * $p < .05$, ~ $p < .10$

3.2.4.2.1 Null model

The null model (Equation 1 below) included no predictors and served to indicate variability at each level. The intra-class correlation coefficient was 5.43/(5.43+0.78+2.79) = 0.603, indicating 60.3% of the total MF variability was due to individual differences (Level-3). In addition, 8.7% (0.78/(5.43+0.78+2.79)) of MF variance could be attributed to between-day differences (Level-2), and 31.0% (2.79/(5.43+0.78+2.79)) to within-day differences (Level-1).

3.2.4.2.2 Growth models

The deviance statistic (difference in -2 Log Likelihood values) between growth models was used to test best-fit. Deviance statistics were tested as chi-squared values with degrees of freedom equal to the number of additional parameters needed to estimate the later model. The random quadratic model (model 4) provided a better fit that the random linear model (model 2) (deviance statistic = 6407.83 – 6356.24 = 51.59, df = 6, p < .001). However, there was no statistically significant variability in quadratic term variance or covariance at Level-3, or the linear intercept covariance at Level-2 (see Model 4 column, Table 6). These model components were discarded in favour of a more parsimonious model by eliminating the quadratic random effect at Level-3 and reverting to a *diagonal* variance-covariance matrix at Level-3, fixing the covariance at 0 (Model 5, Equation 2). The growth curve parameters in model 5 brought a decline in residual variance of 1.26 (2.79 to 1.53), suggesting that 45.2% of within-day fatigue variability could be attributed to growth components (time).

3.2.4.2.3 Covariates

Potential covariates were chosen based on theory and entered in a stepwise manner as recommended by Hox (2010). First, Level-1 covariates were entered (physical exertion, napping, smoking) retaining only those statistically significant. Next, Level-2 covariates (sleep quality, sleep length) and, finally, Level-3 covariates (age, gender, menstrual phase) were entered and retained using the same statistical significance criterion. Statistical significance criterion for covariates was $\alpha < .10$. All Level-1 and Level-2 covariates were person-centred (*assessment - person-mean assessment*)

meaning all estimated within-person effects were free from between-person variance (J. E. Schwartz & Stone, 1998) and model intercept represented fatigue at the person-average level for all covariates (at 1000h).

Only physical exertion (Level-1) and sleep quality (Level-2) were retained as model covariates (model 6, equation 3) which led to a further 0.6% reduction in Level-1 variability and 3.1% reduction in Level-2 variability. Model 6 had better fit than model 5 (deviance statistic = 117.50, df = 2, $p < .001$) and therefore used for hypothesis testing. Group differences (RRMS, control) in diurnal fatigue patterns were examined by entering group as a binary fixed effect at Level-3 with group by linear and group by quadratic interaction effects. The interaction effects described the difference between groups. As a supplementary analysis, the RRMS group was split into those with (RRMS-f) and without (RRMS-nf) "clinically meaningful fatigue" based on the ≥4 FS cut-off, and diurnal fatigue patterns were re-examined between three groups by entering three sets of binary variables.

Equation 1

Model 1: Unconditional growth model

$$FATIGUE_{adi} = \gamma_{000} + V_{00i} + U_{0di} + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$ and $V_{00i} \sim N(0, \sigma_{00}^2)$

Equation 2

Model 5: Random quadratic growth model with quadratic effect fixed at Level-3

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + U_{1di}(TIME_{adi}) + U_{2di}(TIME_{adi})^2 + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \\ U_{2di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$ and

$\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & 0 \\ 0 & \sigma_{10}^2 \end{bmatrix} \right)$

Equation 3

Model 6: Random quadratic covariate growth model with quadratic effect fixed at Level-3

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + \gamma_{300}(EXERTION) + \gamma_{010}(SLEEP\ QUALITY) + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + U_{1di}(TIME_{adi}) + U_{2di}(TIME_{adi})^2 + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \\ U_{2di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix}\right)$ and $\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & 0 \\ 0 & \sigma_{10}^2 \end{bmatrix}\right)$

Where $FATIGUE_{adi}$ is the value of MF for person i on day d at assessment a; $TIME_{adi}$ the time of day (centred at 1000h) for person i on day d at assessment a. Fixed effects are denoted by γ, with γ_{000} indicating the average intercept, γ_{100} the average effect of *TIME*, γ_{200} the average effect of *TIME*² (quadratic). Random effects at Level–3 (individual) are indicated by V, with V_{00i} denoting the deviation of intercept for person i from the average intercept, V_{10i} denoting the deviation of the effect of person i from the average effect of *TIME*, and V_{20i} denoting the deviation of the effect of person i from the average effect of *TIME*². Random effects at Level–2 (day) are indicated by U, with U_{0di} denoting the deviation of intercept for day d from the average intercept of person i, U_{1di} denoting the deviation of the effect of day d from the average effect of *TIME* of person i, and U_{2di} denoting the deviation of the effect of day d from the average effect of *TIME*² of person i. R_{adi} indicates residuals at the level of the individual assessments.

3.2.4.2.4 Contextual effects of stress and mood

Model 6 was used in testing the effects of stressors and mood on MF. Stressor challenge, stressor hindrance, distress, and positive mood scores were all person-centred: fixed effect parameters represented the change in MF associated with a one-unit increase in the predictor above the person-mean. One model was used to test the effects of stressors (i.e., challenge and hindrance entered together) on MF, and another to test associations with mood (distress and positive mood).

To elucidate the direction of effect between mood and fatigue, lagged-effects models were computed whereby each predictor (measured at time1) was entered as a fixed effect into a model of time2 MF (one assessment later), with time1 MF entered as a fixed effect to model pure change in fatigue from time1 to time2. Level-1 covariates and time parameters were lagged (time2). The reverse was also conducted, with lagged distress and positive mood (time2) as outcomes in separate models, with mood (time1) and MF (time1) as predictors.

Using lagged outcomes led to *k*-1 assessments, leaving five assessments per day. Due to model convergence problems with this reduced dataset, a 2-level model was used (assessments within individuals). Although between-day variability in MF was shown to be relatively small (8.7% of total variability), using a 2-level model means results should be taken with caution.

3.2.4.3 Recalled fatigue discrepancy and heuristic bias

The accuracy of EOD fatigue recall was defined as the extent to which they corresponded with average daily MF ratings (*Discrepancy = EOD fatigue – daily mean MF*) (Schneider et al., 2011). Momentary assessments were assumed to provide a random sample of all possible reports (Broderick et al., 2008) which was appropriate given the quasi-random design. Discrepancy was modelled using MLM with person-centred fatigue variability (daily MF *SD* minus person-mean MF *SD*) as a fixed effect with group effect and group*fatigue-variability interaction. Practise effects were tested using a 2x4 factorial ANOVA with group (RRMS or control) and day (1-4) entered as predictors of recall discrepancy.

Peak fatigue was defined as the highest daily MF score, and end fatigue was the A6 event MF rating. To reduce multicollinearity (high inter-correlations) of mean, peak, and end MF ratings, both peak and end were centred on the mean. This method has been used previously in an examination of peak and end effects in pain and fatigue reports (Schneider et al., 2011) and is "consistent with the theoretical concept that peak and end can only introduce bias in recall to the extent that they differ from the actual average of symptoms experienced" (p. 230).

A 2-level multilevel model was built with EOD fatigue as outcome, with four EOD ratings (one per day) nested within individuals:

Equation 4

Null model of EOD fatigue

$$EOD\ fatigue_{di} = \gamma_{00} + U_{0i} + R_{di}$$

Where $R_{di} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$

Where $EOD\ fatigue_{di}$ is the value of EOD fatigue rating for person i on day d. The average intercept is indicated by γ_{00}. Random effects at Level–2 (individual) are indicated by U with U_{0i} denoting the deviation of the intercept for person i from the average intercept. R_{di} indicates the residuals for each unique assessment.

With each Level-1 unit (day) equally distributed in time, different error variance-covariance matrices were examined to determine the model of best-fit, as suggested by Singer and Willett (2003). The AIC and BIC were used to differentiate by goodness of fit (see Appendix K). The *Toeplitz* matrix was chosen.

EOD ratings were viewed as biased by peak and end heuristics if peak and/or end fatigue predicted EOD fatigue after controlling for daily mean fatigue. Therefore daily mean, peak, and end fatigue were entered as fixed effects permitted to vary (random effect) between individuals, if there was significant variability to model. Group (RRMS, control) was entered as a fixed effect at Level-2 with group*mean, group*peak, and group*end interaction effects. The interaction effects determined group differences in the uses of peak and end heuristics.

3.3 Results

Analysis was based on 1661 completed events (Level-1) nested within 304 days (Level-2) nested within 76 participants (Level-3). Compliance with the EMA protocol (A1-A6) was excellent in both groups, with 91.1% of scheduled quasi-random events completed (1661 out of 1824). The lowest individual compliance rate was 62.5% met by three participants (one RRMS, two control) and the highest was 100%, achieved by 21 participants (10 RRMS, 11 control). Residual distributions for each multilevel model did not substantially deviate from normality (residual histograms presented in Appendix L).

Table 7 Group Differences for Study Fatigue Measures

	RRMS		Control		
	Mean (*SD*)	Min, Max	Mean (*SD*)	Min, Max	*p*
Baseline measures					
FS-Total	17.58 (7.09)	1, 32	11.55 (2.87)	4, 21	<.001
FS-Physical	11.18 (4.89)	0, 20	7.26 (2.34)	2, 16	<.001
FS-Mental	6.39 (2.66)	1, 12	4.29 (0.96)	2, 7	<.001
MFIS	9.39 (4.25)	0, 16	3.87 (3.82)	0, 14	<.001
EOD measures (person-mean)					
EOD Fatigue	5.79 (2.73)	0.00, 10.00	2.05 (1.96)	0.00, 7.25	<.001
Momentary measures (person-mean)					
MF	5.08 (2.29)	0.43, 8.87	2.43 (1.71)	0.00, 6.79	<.001

Note. Due to non-normal distributions in all fatigue measures, Mann-Whitney U tests were conducted for all group comparisons. FS indicates Fatigue Scale; MFIS, Modified Fatigue Impact Scale, EOD, End-of-day; MF, momentary fatigue.

3.3.1 Fatigue reports

Table 7 reports group differences for fatigue measures (baseline, EOD, momentary). As expected, baseline fatigue severity was higher in the RRMS group than control group, d = 1.11, 95% CI [0.62, 1.58] (see Figure 3). Greater baseline fatigue severity in RRMS was reflected in physical fatigue, d = 1.02, 95% CI [0.53, 1.49], and mental fatigue subscales, d = 1.05, 95% CI [0.56, 1.52]. Baseline fatigue impact was also higher in the RRMS group, d = 1.36,

95% CI [0.85, 1.85] (see Figure 3). Similar, large, effect sizes were observed for person-mean EOD fatigue severity reports, d = 1.57, 95% CI [1.04, 2.07], and person-mean MF ratings, d = 1.32, 95% CI [0.81, 1.80]. The RRMS group did not differ from the control group for average sleep quality, U = 691.00, p = .747, or sleep duration, t = 1.26, df = 74, p = .214, 95% CI [-0.15, 0.67].

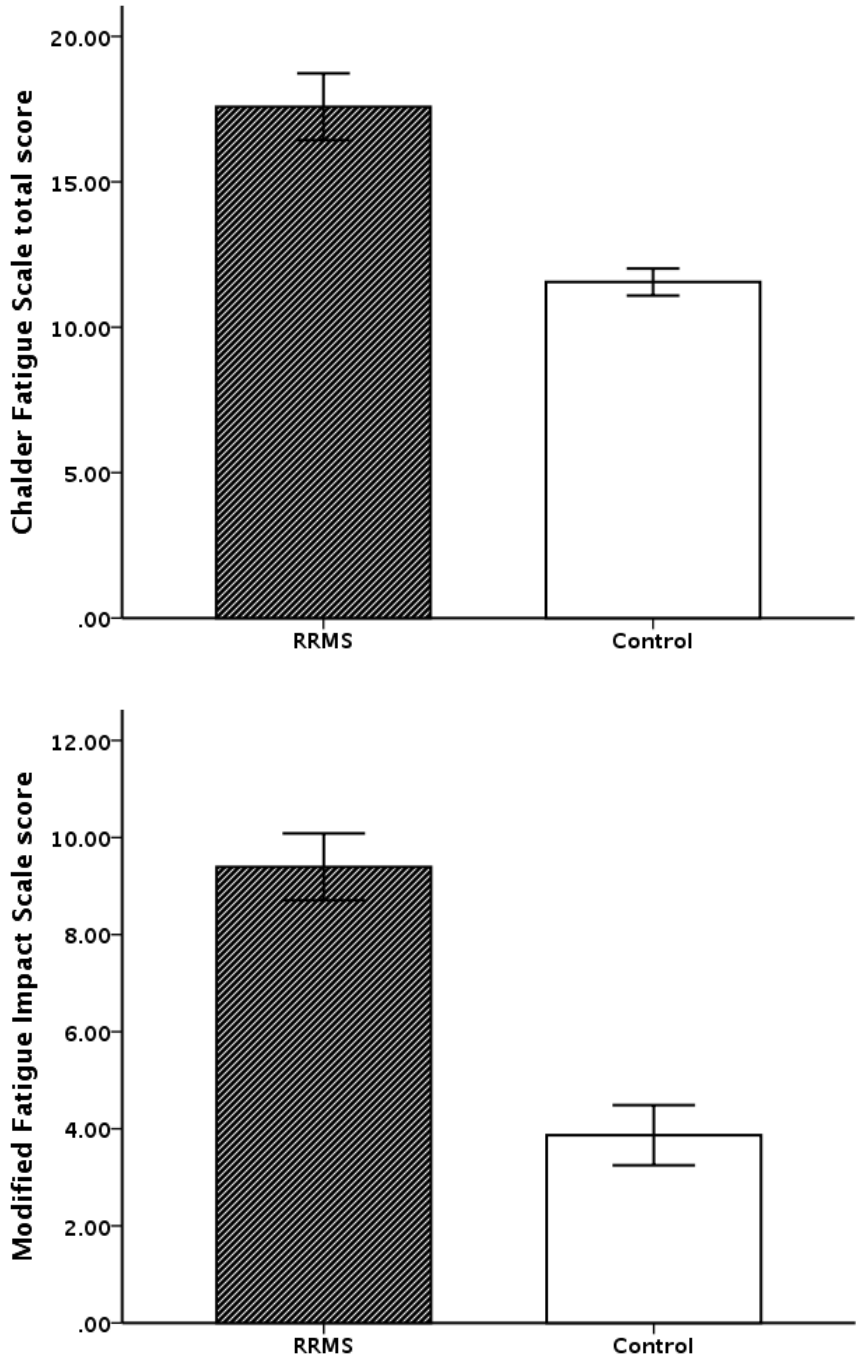


Figure 3. Group differences in baseline fatigue severity and impact.

3.3.2 Variability in fatigue experience

The partitioning of variance in the null model revealed there was significant variability both within-days, σ_R^2 = 2.79, *SE* = 0.10, $p < .001$, and between-days, σ_0^2 = 0.78, *SE* = 0.13, $p < .001$. The model attributed 31.0% of total MF variance within-days, 8.7% between-days, and the remaining 60.3% between-individuals. Figure 4 demonstrates within-person variability in MF ratings in nine randomly-selected individuals from the RRMS group. It was evident that some individuals reported fairly stable fatigue experience (e.g., ID1 and ID72) whereas others had variable fatigue experiences (e.g., ID42 and ID83). Figure 4 highlights the loss of within-person information when summarised fatigue reports are used in research or clinical practise; ID42 and ID54 reported very similar average daily MF (M_f = 4.14 and 4.63, respectively) yet their fatigue experience in everyday life is dissimilar.

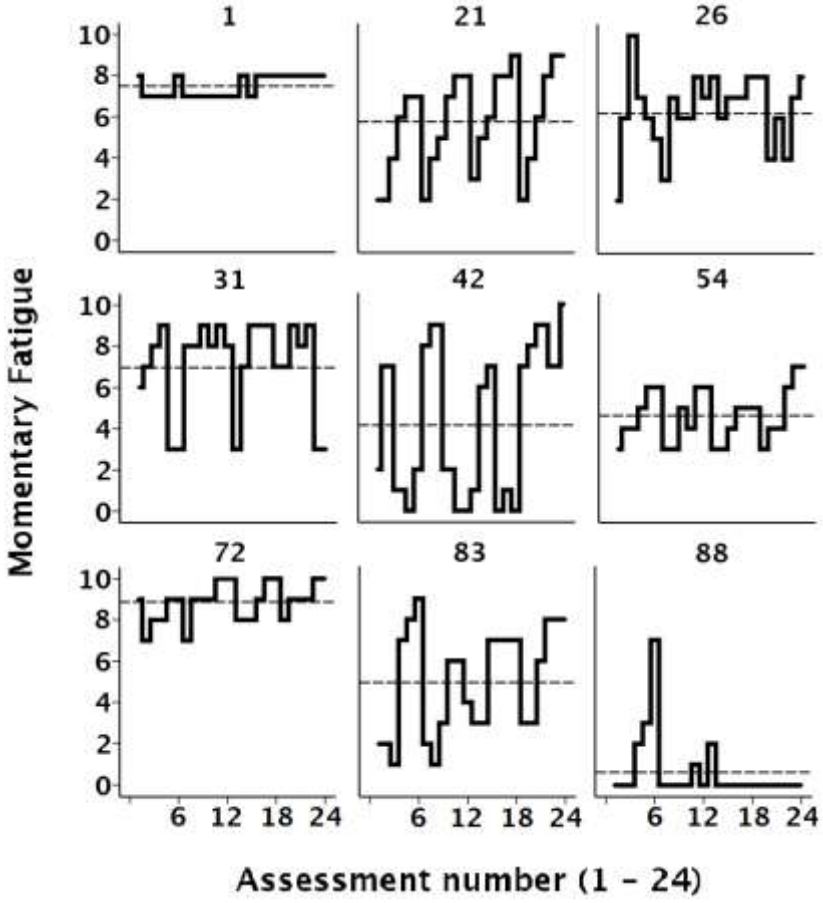


Figure 4. Step line charts of momentary fatigue ratings for nine randomly-selected individuals from the RRMS group.

Note. Assessment numbers 6, 12, 18, and 24 denote final assessments from each day. Dashed line indicates person-mean momentary fatigue.

Participant *SD*s were no different between groups, $U = 699.00$, $p = .81$. In the RRMS group, 92.5% of MF reports were above 0, compared to 69.1% in controls, $U = 330.50$, $p < .001$. MF scores above 5 (scale mid-point) were reported in 53.3% of RRMS assessments, and 17.4% of control group assessments, $U = 321.50$, $p < .001$. The maximum MF rating (10) was rarely used in either group (2.2% RRMS, 0.1% control).

3.3.3 Typical diurnal fatigue patterns

Table 8 presents parameter estimates and significance levels for the diurnal fatigue patterns in both groups. MF ratings were, on average, 2.31 units higher in the RRMS group than control group at 1000h, $p < .001$, 95% CI [1.28, 3.35]. Average diurnal fatigue patterns showed MF increased over the day in both groups, but with different characteristics. The control group demonstrated a robust linear trajectory whereby fatigue increased by 0.29 units every hour, on average, $p = .010$, 95% CI [0.07, 0.51]. However, the RRMS group fatigue trajectory had a statistically significant quadratic component such that the trajectory could be modelled by an increase of 0.51 units per hour, $p < .001$, 95% CI [0.29, 0.73], and a coinciding decrease of $0.03*time^2$ units, $p = .009$, 95% CI [-0.05, -0.01], representing a decelerating diurnal fatigue pattern (see Figure 5). In a sensitivity analysis including depressive symptoms as a covariate, average 1000h MF was somewhat reduced (14% reduction in RRMS group; 24% reduction in control group), but fatigue trajectories remained unchanged. When paid employment (full-time, part-time, or unemployed) was added as a potential moderator, there was no effect of employment on average MF at 1000h in either group (*p*s > .73), or on linear (*p*s > .18) or quadratic parameters (*p*s > .20).

Table 8 Multilevel Parameter Estimates for Effects of time and time^2 on Momentary Fatigue Ratings in People with RRMS and Healthy Participants.

		RRMS group	Control group	Difference between groups
Model 6				
Overall level (intercept)	γ (*SE*)	3.40 (0.37)***	1.08 (0.37)**	2.31 (0.52)***
Linear growth (time)	γ (*SE*)	0.51 (0.11)***	0.29 (0.11)*	0.22 (0.16)
Quadratic growth (time^2)	γ (*SE*)	-0.03 (0.01)**	-0.004 (0.01)	-0.02 (0.02)˜
Model 6 with depressive symptoms moderator				
Overall level (intercept)	γ (*SE*)	2.91 (0.53)***	0.83 (0.42)*	2.08 (0.55)***
Linear growth (time)	γ (*SE*)	0.51 (0.11)***	0.29 (0.11)*	0.22 (0.16)
Quadratic growth (time^2)	γ (*SE*)	-0.03 (0.01)**	-0.003 (0.01)	-0.02 (0.02)˜

Note. The fixed effects of covariates and the random effects of intercept, time, and time^2 are not shown. *** $p < .001$, ** $p < .01$, * $p < .05$, ˜ $p = .10$.

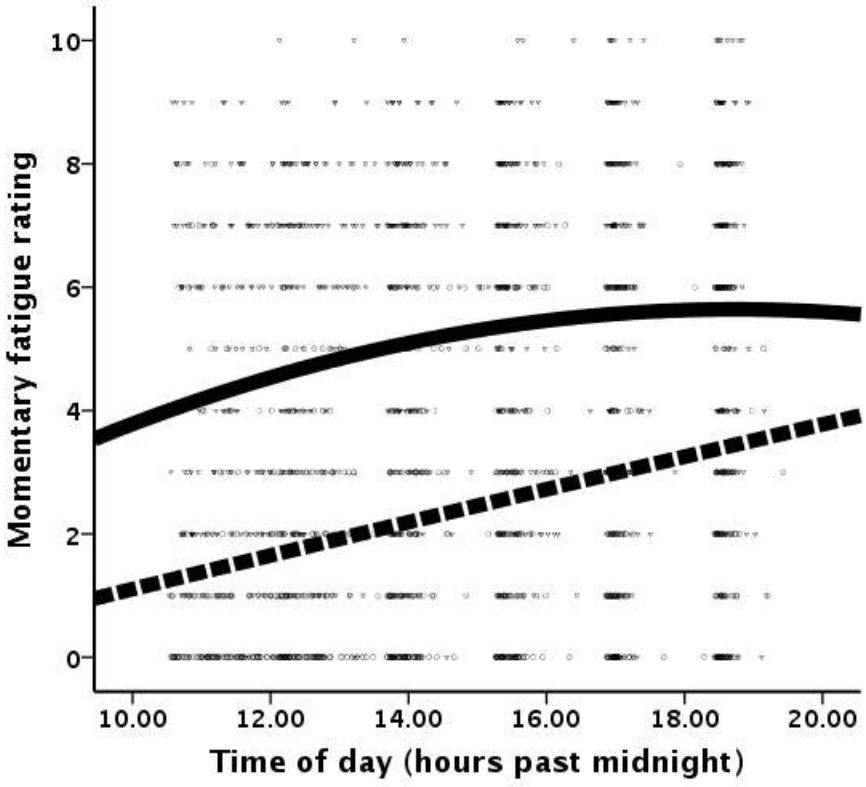


Figure 5. Average fatigue diurnal patterns in the RRMS (solid line) and control (dashed line) groups.

Note. Triangles (RRMS) and circles (control) represent unique momentary fatigue ratings. Fatigue trajectories depicted in figure do not account for sleep quality or physical exertion.

3.3.4 Clinically meaningful fatigue in RRMS

The RRMS group was split into those with (RRMS-f, n = 21) and without (RRMS-nf, n = 17) "clinically meaningful fatigue" by ≥4 FS cut-off. Average person-mean MF was higher in the RRMS-f group (M_f = 6.12 ±1.88) than the RRMS-nf group (M_f = 3.81 ±2.14; U = 74.50, p = .002). Table 9 presents group comparisons for key variables; there were no statistically significant differences for employment, average sleep quality, or gender. Although the RRMS-nf subgroup was slightly older than the RRMS-f and control groups, this was not a statistically significant difference. The RRMS-f subgroup was more disabled than the RRMS-nf subgroup, but not to statistical significance. There was no difference on the FS between RRMS-nf and healthy controls, but the RRMS-nf group did score higher on the MFIS-5, U = 186.50, p = .012.

Table 9 Demographic and Clinical Characteristics of RRMS-f, RRMS-nf, and Control Groups

	RRMS-f	RRMS-nf	Control	*p*
n	21	17	38	
Age	40.38 (6.94)	43.76 (8.00)	40.34 (8.15)	.27
Gender	18f, 3m	13f, 4m	31f, 7m	.77
Paid employment	10 full-time, 6 part-time, 5 unemployed	10 full-time, 4 part-time, 3 unemployed	24 full-time[a], 8 part-time, 5 unemployed	.78
Sleep quality	5.94 (1.81)	6.27 (1.24)	6.23 (1.99)	.86
Sleep duration (hours)	7.96 (1.10)	7.67 (0.82)	7.57 (0.80)	.29
EDSS	4.69 (0.89)	3.79 (1.71)		.06
FS	22.71 (4.84)	11.24 (3.11)	11.55 (2.87)	<.001
MFIS	11.71 (3.21)	6.53 (3.62)	3.87 (3.82)	<.001

Note. Mean (*SD*) shown. EDSS indicates Expanded Disability Status Scale; FS, Fatigue Scale; MFIS, Modified Fatigue Impact Scale. Kruskal-Wallis tests performed to check group differences, except Sleep Quality where ANOVA was conducted and Chi Square test for Gender and Paid Employment variables.

[a] One individual did not respond to question regarding working hours; however, they did indicate they were employed.

3.3.4.1 Diurnal fatigue pattern

MF was, on average, 2.62 units higher in the RRMS-f subgroup than the RRMS-nf subgroup at 1000h, $p < .001$, 95% CI [1.25, 3.96], and 3.49 units higher than the control group, $p < .001$, 95% CI [2.38, 4.63] (see Table 10). The RRMS-nf subgroup was, on average, 0.88 units higher at 1000h than the control group, but this was not statistically significant, $p = .151$, 95% CI [-0.33, 2.09]. A similar statistically significant positive linear growth trend was present in both RRMS-f, $p = .002$, 95% CI [0.18, 0.78], and RRMS-nf subgroups, $p = .002$, 95% CI [0.20, 0.86], and there was a negative quadratic component which was not statistically significant (probably due to smaller *n*) in both RRMS-f, $p = .052$, 95% CI [-0.06, 0.0002], and RRMS-nf subgroups, $p = .102$, 95% CI [-0.06, 0.005]. Figure 6 suggests quadratic diurnal fatigue patterns present in both RRMS subgroups, yet missing in the control group, and demonstrates the variability in diurnal fatigue patterns even within groups. In a sensitivity analysis, depressive symptoms (all $ps > .14$) and employment status (all $ps > .28$) did not significantly change any fixed effect or interaction parameters in any group.

Table 10 Multilevel Model Parameters of Group Differences in Average Fatigue Levels at 1000h and Diurnal Fatigue Patterns

	γ (*SE*)	Group comparisons	
RRMS-f		vs RRMS-nf	vs Control
Overall level (intercept)	4.59 (0.46)***	2.62 (0.69)***	3.49 (0.57)***
Linear growth (time)	0.48 (0.15)**	-0.06 (0.23)	0.20 (0.19)
Quadratic growth (time^2)	-0.03 (0.01)~	0.002 (0.01)	-0.03 (0.02)
RRMS-nf		vs Control	
Overall level (intercept)	1.97 (0.46)***	0.88 (0.61)	
Linear growth (time)	0.53 (0.17)**	0.25 (0.20)*	
Quadratic growth (time^2)	-0.03 (0.02)~	-0.02 (0.02)	
Controls			
Overall level (intercept)	1.08 (0.34)**		
Linear growth (time)	0.29 (0.11)*		
Quadratic growth (time^2)	-0.004 (0.01)		

*** $p < .001$, ** $p < .01$, * $p < .05$, ~ $p \leq .10$.

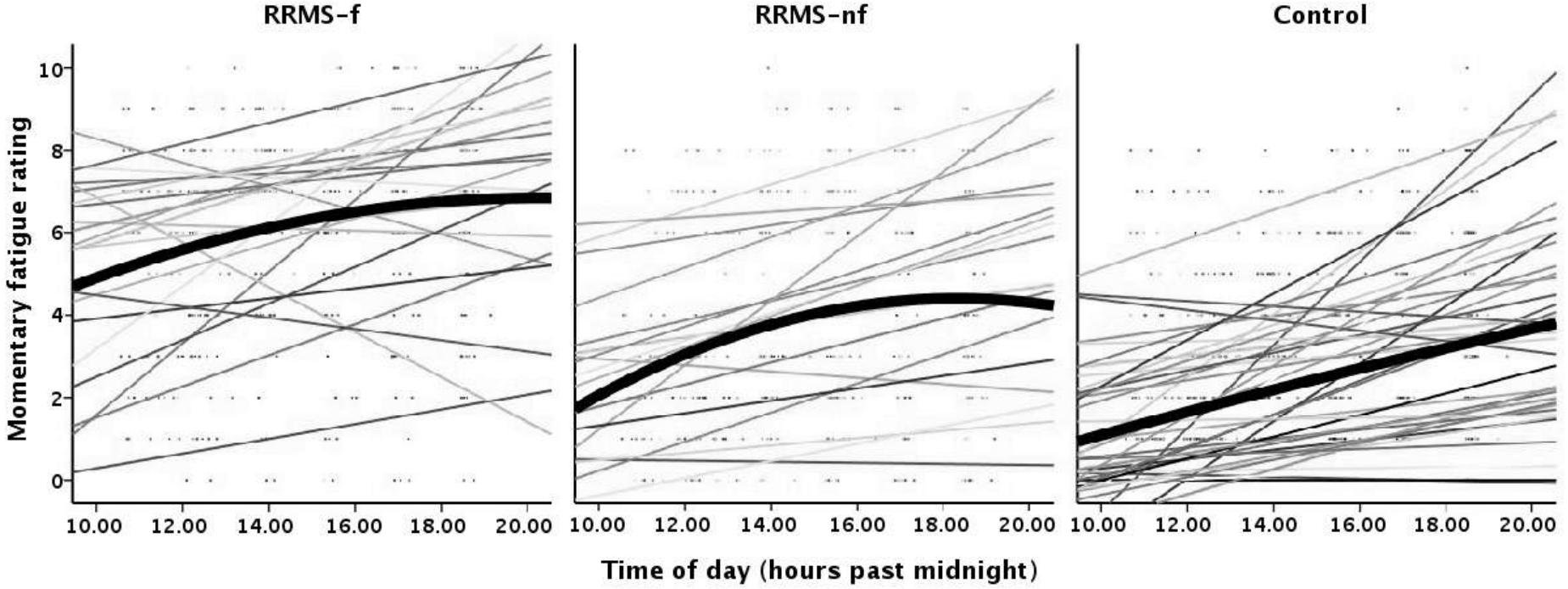


Figure 6. Spaghetti plots of average group (thick) and individual-specific (thin) diurnal fatigue patterns as a function of time of day.

Note. Individual-specific slopes are presented with linear component only to demonstrate variability; group slopes by average quadratic trend. Small markers indicate individual momentary fatigue assessments. Fatigue trajectories do not account for daily sleep quality or momentary physical exertion.

3.3.5 Contextual effects of stress and mood

There were no group differences for within-person means of stressor challenge (RRMS, M = 4.74 ±1.40; control M = 4.58 ±1.74, $p > .05$) or stressor hindrance (RRMS, M = 1.95 ±1.26; control M = 1.46 ±1.11, $p > .05$). However, both distress (RRMS, M = 1.93 ±1.17; control, M = 1.28 ±1.32, $p < .01$) and positive mood (RRMS, M = 4.84 ±1.54; control, M = 5.74 ±1.74, $p < .05$) differed significantly between groups. Diurnal patterns of stressor and mood items in both groups are presented for reference in Appendix M.

The diurnal fatigue pattern accounted for 45.2% of MF within-day variability, but left a substantial 54.8% of within-day variability unexplained. The contextual effects of stressor challenge and stressor hindrance on MF in each group are detailed in Table 11. Prior challenge was not a statistically significant predictor of MF in either the RRMS group, p = .18, 95% CI [-0.02, 0.10], or control group, p = .84, 95% CI [-0.07, 0.05]. However, for every one-unit increase in prior hindrance above a person's usual level, there was an associated 0.18 increase in MF in the RRMS group, $p < .001$, 95% CI [0.09, 0.26]. A somewhat larger effect was present in the control group, $p < .001$, 95% CI [0.25, 0.34], but a group interaction did not reach statistical significance, p = .10, 95% CI [-0.02, 0.23]. Level-1 residual variance, reflecting within-day variability, was reduced by 8% (1.616 to 1.487) with the addition of stressor hindrance as a fixed effect in the growth model, and Level-2 variance reduced by 13% (2.829 to 2.449), representing an overall 11% reduction in within-subject fatigue variability.

Table 11 Multilevel Estimates for Within-Subjects Effects of Daily Life Stressors on Fatigue

	RRMS	Control	Group difference
	γ (*SE*)	γ (*SE*)	γ (*SE*)
Stressor Challenge	0.04 (0.03)	-0.006 (0.03)	0.05 (0.04)
Stressor Hindrance	0.18 (0.04)***	0.29 (0.05)***	-0.11 (0.07)˜

Note. Fixed effects of covariates and the fixed and random effects of intercept, time and time^2 are not shown.

*** $p < .001$, ** $p < .01$, * $p < .05$, ˜$p = .10$.

The contextual associations of momentary distress and positive mood with MF are detailed in the model outputs reported in Table 12; both predictors were associated with MF in both groups. In RRMS, a one-unit increase in distress beyond a person's usual experience was associated with a 0.17 increase in fatigue, $p < .001$, 95% CI [0.08, 0.26]. A one-unit increase in positive mood above usual experience was associated with a 0.39 decrease in fatigue in the RRMS group, $p < .001$, 95% CI [-0.48, -0.30]. Both groups showed similar momentary associations of mood with fatigue, and there was no interaction effects for distress*group, $p = .36$, 95% CI [-0.07, 0.20], or positive mood*group, $p = 0.59$, 95% CI [-0.10, 0.17]. Level-1 residual variance was reduced by 10% (1.616 to 1.452), and Level-2 variance by 16% (2.829 to 2.367); a total of 14% of within-individual fatigue variability accounted for by mood fluctuations.

Table 12 Multilevel Estimates for Within-Subjects Associations of Mood with Fatigue

	RRMS	Control	Group difference
	γ (*SE*)	γ (*SE*)	γ (*SE*)
Momentary Distress	0.17 (0.05)***	0.24 (0.05)***	-0.07 (0.07)
Momentary Positive Mood	-0.39 (0.04)***	-0.35 (0.05)***	-0.04 (0.07)

Note. Fixed effects of covariates and the fixed and random effects of intercept, time and time^2 are not shown. *** $p < .001$, ** $p < .01$, * $p < .05$.

Associations with fatigue for any stress or mood predictor did not differ between RRMS-f and RRMS-nf groups to statistical significance (all *p*s >.05). Lagged-effects models examining the temporal relationship between mood and fatigue did not demonstrate any causal relationships extending beyond a momentary association (see Table 13).

Table 13 Multilevel Estimates for Lagged Model testing the Direction of Within-Subjects Effects of Mood on Fatigue

	RRMS	Control	Group difference
	γ (*SE*)	γ (*SE*)	γ (*SE*)
Change in Fatigue (Lagged Fatigue)			
Distress	0.03 (0.05)	0.002 (0.06)	0.03 (0.07)
Positive Mood	-0.05 (0.05)	0.01 (0.06)	-0.07 (0.08)
Change in Distress (Lagged Distress)			
Fatigue	-0.01 (0.04)	0.01 (0.03)	0.01 (0.05)
Change in Positive Mood (Lagged Positive Mood)			
Fatigue	-0.02 (0.03)	0.01 (0.03)	-0.03 (0.05)

Note. Fixed effects of covariates (including prior outcome) and the fixed and random effects of intercept, time and time^2 are not shown. *p < .05.

3.3.6 End-of-day fatigue recall discrepancy

EOD fatigue ratings were provided on 260 of 304 assessment days (85.5%), with similar response rates in both the RRMS group (84.2%) and control group (86.8%). In both groups, Kendall's tau *b* correlations between EOD fatigue and daily mean MF were high on all 4 days, as expected (RRMS: .624 < r < .769; Control: .638 < r < .667). The RRMS group were characterised by an average overestimation in EOD fatigue of 0.73 units (12.4% increase) above mean MF scores, whereas the control group had an average underestimation of 0.38 units (15.6% decrease), representing a statistically significant difference in EOD discrepancy between groups, γ = 1.13, *SE* = 0.25, p < .001, 95% CI [0.65, 1.62]. The distribution of discrepancy scores is depicted in Figure 7. If discrepancy was defined as deviation from 0 in any direction (one-unit overestimation = one-unit underestimation), there was no significant difference between groups in accuracy (p = .104).

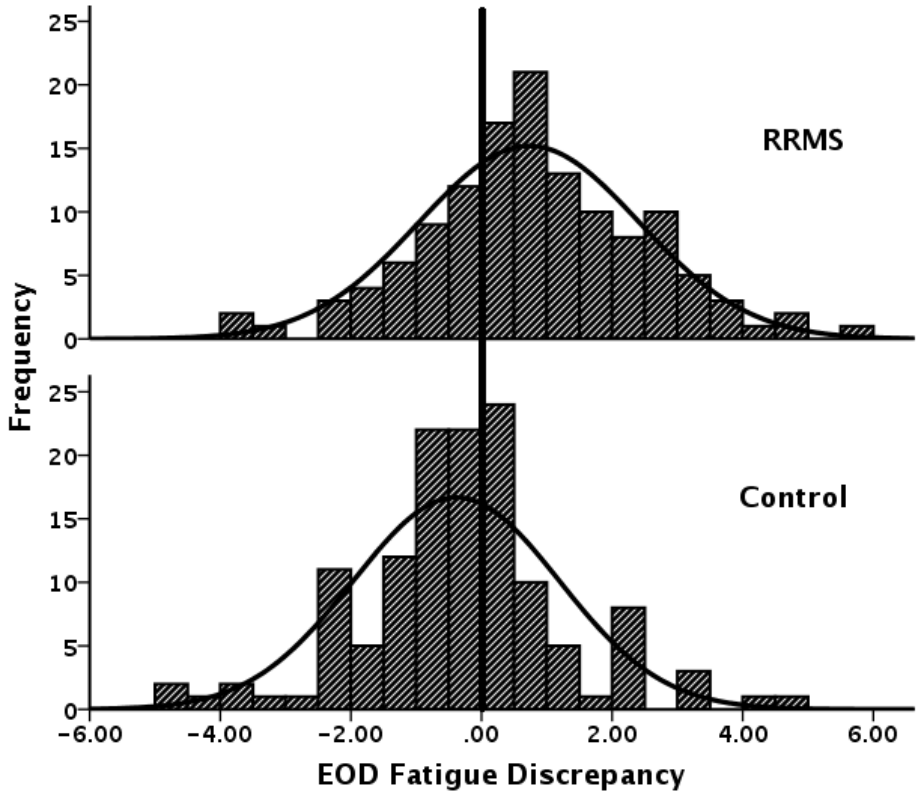


Figure 7. Distribution of EOD fatigue rating discrepancies in each group. Ratings to the left (underestimation) and right (overestimation) of 0 indicate discrepancy.

The within-subjects association of discrepancy with fatigue variability (*SD*) was not quite statistically significant in either the RRMS group, $\gamma = 0.36$, $p = .053$, 95% CI [-0.01, 0.73], or the control group, $\gamma = -0.28$, $p = .070$, 95% CI [-0.59, 0.02], although there was a highly statistically significant difference between groups in the association of variability with discrepancy, $\gamma = 0.65$, $p < .01$, 95% CI [0.17, 1.13]: increases in fatigue variability from an individual's usual fatigue variability were more likely to result in daily fatigue recall overestimation in the RRMS group and underestimation in the control group.

Figure 8 demonstrates consistent overestimation on each day within the RRMS group, and consistent underestimation within the control group. Although there appears to be a potential practise effect evident in the control group, this was not statistically significant in the factorial ANOVA ($p = .395$).

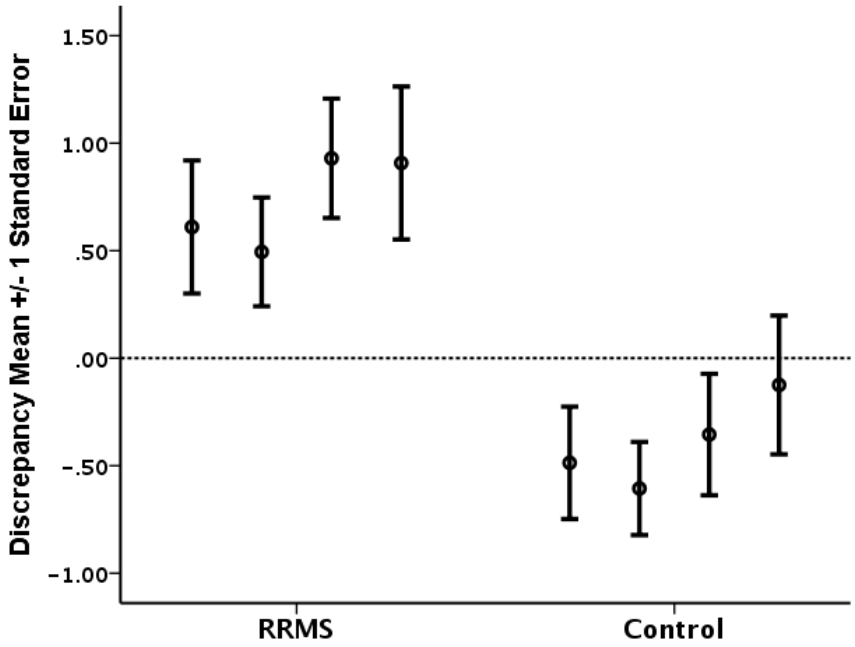


Figure 8. Average EOD fatigue discrepancies by day. Error bars indicate standardised error of the mean, clustered by day (1-4, left to right).

3.3.6.1 Peak and end effects

Peak MF ratings were, on average, 1.71 units higher (*SD* = 1.13) than the person-mean in the RRMS group, and 1.81 units higher (*SD* = 1.25) in the control group. End MF ratings were, on average, 0.58 units higher (*SD* = 1.59) than the person-mean in the RRMS group, and 0.94 units higher (*SD* = 1.61) in the control group. Table 14 presents results of the within-subjects analysis of EOD fatigue ratings. Across all participants, 65.01% of within-subject variance in EOD fatigue was accounted for by daily mean MF. There was a statistically significant association of mean fatigue with EOD fatigue in both the RRMS group, $p < .001$, 95% CI [0.36, 0.93], and control group, $p < .001$, 95% CI [0.68, 1.22]. Although this association was seemingly stronger in the control group, there was no statistically significant difference between groups, $p = .134$, 95% CI [-0.68, 0.10]. When daily peak and end fatigue were added to the model, an additional 1.86% of within-person variability was accounted for. In this second step, peak fatigue was relevant, although not to statistical significance, only in the RRMS group, $p = .082$, 95% CI [-0.04, 0.65], and not the control group, $p = .312$, 95% CI [-0.42, 0.13]. Similarly, end fatigue was

relevant only within the RRMS group, p = .026, 95% CI [0.03, 0.39], and not the control group, p = .600, 95% CI [-0.27, 0.16]. This represented a statistically significant difference between groups for use of the peak heuristic, p = .047, 95% CI [0.01, 0.89], and a trend towards a difference between groups for use of the end heuristic, p = .066, 95% CI [-0.02, 0.54].

Table 14 Multilevel Parameters Predicting End of Day Fatigue Ratings from Daily Mean, Peak, and End Momentary Fatigue Scores

	RRMS	Control	Group comparison	Variance Explained	Additional Variance explained
	γ (*SE*)	γ (*SE*)	γ (*SE*)		
Step 1				*65.01%*	
MF mean	0.65 (0.14)***	0.95 (0.13)***	-0.21 (0.19)		
Step 2				*66.87%*	*1.86%*
MF peak	0.30 (0.17)˜	-0.14 (0.14)	0.45 (0.22)*		
MF end	0.21 (0.09)*	-0.06 (0.11)	0.27 (0.14)˜		

Note: MF indicates Momentary Fatigue rating; mean, daily mean; peak, highest within-day MF rating; end, MF at last momentary assessment event (A6).
*** p < .001, * p < .05, ˜ p < .10

3.3.7 Results summary

Diurnal fatigue patterns in the RRMS group retained a negative quadratic growth component, indicating people with RRMS are more fatigable earlier in the day than later; healthy individuals retained a typical linear diurnal fatigue pattern. Stressor hindrance was associated with fatigue within-persons in both groups. Whilst there was a momentary association between mood and fatigue in both groups, models were unable to identify the causal direction of this association.

EOD recall biases in fatigue reports demonstrated an overestimation in the RRMS group, with both most-intense fatigue (peak) and recent fatigue (end) contributing to EOD reports above and beyond mean MF. The control group demonstrated an underestimation of fatigue experience in EOD reports,

and there was no evidence of any contribution of peak or end heuristics in EOD report provision.

3.4 Discussion

Despite its high prevalence, MS fatigue experience is poorly understood. This chapter demonstrated elevations in average fatigue severity in pwRRMS compared to healthy individuals in everyday life, while a significant proportion of within-day variability in fatigue was characterised by increasing (but decelerating) diurnal fatigue patterns in RRMS. Stressors hindering goal achievement (hindrance) were associated with increased fatigue in all groups, whereas those stressors representing challenge had no effect on fatigue. Momentary associations of fatigue with increased distress and reduced positive mood were present in all populations, but it was not possible to make any causative inferences from the data. Overestimation in recalled EOD fatigue ratings were demonstrated in RRMS, with both peak (primacy) and end (recency) heuristics contributing to EOD reporting after accounting for actual average fatigue severity in RRMS. These biases were not present in EOD fatigue ratings provided by the control group.

3.4.1 Fatigue experience in everyday life and contextual effects

Existing literature, predominantly employing recalled measures of experience, suggested fatigue levels in the RRMS group would be higher than the control group (for example, Freal et al., 1984; Mollaoğlu & Üstün, 2009). Unlike the only other study to examine fatigue in MS in everyday life with an EMA strategy (E. Kim et al., 2010), the present study employed a healthy control comparator group. The present study also had a superior compliance rate (91% versus 64%) probably due to having audible prompts and item response embedded within the same electronic platform. In the present study, average daily MF and EOD fatigue ratings were higher in RRMS than controls with a large effect size. Along with finding clearly elevated fatigue levels in RRMS at 1000h with multilevel growth models, these findings add to the literature by demonstrating elevated fatigue levels in an RRMS population in everyday life.

The diurnal fatigue pattern in RRMS was expected to show greater increases over time than healthy controls, but there was limited evidence to

support this. Although fatigue levels typically increased quicker in the RRMS group in the early part of the day, this increase decelerated such that a drawn-out peaking of fatigue was evident in the late afternoon and early evening (between 1600h to 1900h). The RRMS group's quadratic growth component was present in those with and without "clinically meaningful fatigue". The typical diurnal fatigue pattern for the healthy control group was a consistent positive linear trajectory with no obvious peak or deceleration within the 1000h to 2000h measurement window. Individuals in the RRMS-nf group demonstrated the decelerating diurnal fatigue pattern despite a typical 1000h fatigue level similar to healthy controls. These findings support early retrospective descriptions of late-afternoon peaks in fatigue severity in MS (Freal et al., 1984) and corroborate observations of decelerating patterns in lab-based EMA studies in MS (Feys et al., 2012; Schwid et al., 2003). The present findings indicate not only a quantitative difference in the experience of fatigue in RRMS, but also a qualitative one: whether experiencing severe fatigue or not, pwRRMS appear more susceptible to increases in fatigue (fatigability) earlier in the day than later.

Only one previous study has examined diurnal fatigue patterns in a clinical group (breast cancer survivors) and healthy control group (Curran et al., 2004). Here, fatigue was also measured using a 0-10 VAS with end points "no fatigue" and "worst possible fatigue". The study found no discernable differences in the diurnal fatigue pattern between groups, unlike here, but did find a similar higher fatigue level at baseline and all other time points in the clinical group. The study by Curran et al. (2004) may have lacked the precision to observe decelerations in fatigue trajectories due to a limited number of momentary assessments, and highlights a strength of the present study: quasi-random assessments ensured a spread of assessments throughout the day.

Conceptualisations of fatigue are likely to be different for individuals who have fatigue which is symptomatic of a clinical disorder than those who are healthy. It is an advantage of within-subjects analyses that the individual is their own control, but it is clear that what individuals consider "extreme fatigue" would likely be calibrated by past experience. There have been several qualitative explorations of the fatigue experience between clinical groups and healthy individuals (Crow & Hammond, 1996; Gledhill, 2005) which have

highlighted vastly different fatigue experiences. Results should be interpreted with this in mind.

The final growth model of fatigue, while accounting for 45% of variance, left a substantial proportion of unexplained variability. The study here demonstrated that periods of stress elicit a corresponding increase in fatigue, while controlling for growth patterns. This finding builds on previous work demonstrating an association between stress and fatigue between individuals (R. F. Brown et al., 2009; Trojan et al., 2007). However, it was clear from findings that only changes in stressors representing goal hindrance elicited increases in fatigue; increases in challenge stressors did not. Although this pattern of findings was replicated in both pwRRMS and controls, the size of the effect was somewhat smaller in pwRRMS than controls. There is no EMA study with which to compare these findings. However, a study of 696 students using retrospective assessments observed an association between hindrance stressors and exhaustion ($\gamma = 0.27$, $p < .05$) that was almost twice the size of the association between challenge stressors and exhaustion ($\gamma = 0.14$, $p < .05$) in between-subjects analyses (LePine, LePine, & Jackson, 2004). While LePine et al. (2004) suggest both stressor types were associated with fatigue (exhaustion), the study supported a greater size of effect for hindrance stressors which was also the case in the present study.

The present study suggests that avoiding activity is not necessarily an adaptive coping mechanism for pwRRMS with concerns over symptomatic fatigue; unless the activity is appraised as hindering goal achievement, there should be no expectation that increased activity would change fatigue experience. Future research should focus on potentially modifiable factors that can influence the extent to which stressors are appraised as hindrances rather than challenges, possibly via individual differences in goal orientation (LePine et al., 2004).

It was not possible to discern the direction of effect between fatigue and mood from the data. Although increased distress and, particularly, decreased positive mood were associated with fatigue at the point of measurement, mood did not predict future change in fatigue nor did fatigue predict change in mood. This would suggest fatigue and mood to be very closely temporally related, and it would take a far more intensive EMA design (more assessments per day) to observe the direction of effect, if there is not a bidirectional

relationship. The fact neither momentary distress nor momentary positive mood followed a diurnal pattern similar (or the inverse) to fatigue (diurnal patterns presented in Appendix M) suggests mood and fatigue are very different phenomena.

3.4.2 Recalled fatigue and discrepancy

The validity of retrospectively-provided symptom reports is an important issue for both research and clinical practise. The hypothesised overestimation in EOD fatigue reports was only present within the RRMS group; indeed, the healthy control group typically demonstrated underestimation of daily fatigue in EOD reports. It was expected the RRMS group would be more accurate in providing EOD fatigue reports due to being more used to experiencing and reporting fatigue, but there was no evidence of this. There was no difference between groups in report accuracy; however, the RRMS were generally inaccurate by overestimation, whereas the control group were inaccurate by underestimation. This difference seemed partly driven by different relationships of EOD fatigue summaries with within-day fatigue variability. Although increased fatigue variability was associated with higher discrepancy in EOD recall in both groups, the RRMS group tended to overestimate as a result and the control group tended to underestimate.

The 0.71 unit overestimation in EOD fatigue reports in the RRMS group on the 0-10 VAS unlikely represents a difference that is clinically-meaningful. As stated previously, a difference of 18.0-20.0 on a 0-100 VAS scale has been described as minimal for a clinically-significant difference in pain ratings (Grilo et al., 2007; Hägg et al., 2003). If scaled down to the 0-10 VAS scale used here for fatigue severity, the 0.71 unit difference between EOD fatigue and MF is far below the 1.8-2.0 threshold. However, 19.8% of assessment days in the RRMS group surpassed a 1.8 threshold, so sources of this inaccuracy are an important consideration.

It was expected that peak and end biases in EOD fatigue reports would be less prevalent in pwRRMS than healthy individuals due to more experience with the symptom and symptom reporting. However, variability in the EOD rating was attributable to peak and end effects only in the RRMS group. Sixty-five percent of within-subjects EOD fatigue variability was attributable to daily mean MF, which was somewhat higher than the 52% reported by Schneider et

al. (2011) in EOD fatigue ratings from rheumatology patients. However, the present study demonstrated additional variability accounted for by peak and end biases in RRMS, which were not observed by Schneider et al. (2011). Peak and end biases in daily recall assessments have been demonstrated in pain reporting (for example, Jensen et al., 2008; Schneider et al., 2011); however, this is the first study to show these biases in fatigue recall in any population.

These results provide new information about how pwRRMS go about remembering their fatigue, providing insight into the use of heuristics in fatigue recall in everyday life. Researchers and clinicians should be aware that fluctuations in primacy (peak) and recency (end) severity from one day to the next in RRMS can bias the cognitive restructuring of past fatigue experience for recall at the end of the day. Whether this is a cause for concern depends on the purpose of the measurement. A lack of precision in fatigue measurement in research trials, for instance, may result in the unwarranted success or unjustified failure of a treatment or intervention. However, the bias does not seem to equate to a clinically significant problem, and thus the clinician may interpret EOD fatigue reports as summary ratings containing small but significant overrepresentations of peak and recent fatigue (Stone et al., 2000).

3.4.3 Limitations and future directions

There are several limitations to note. Although steps were taken to exclude or model covariates of fatigue, some factors were not possible to control. Data was not collected to clearly define assessment days as working or non-working days and, indeed, to identify when an individual was in work or not. Participants could indicate whether they were "working/studying", but this did not delineate working at home from being in the workplace. It would have been informative to know the working hours for each participant's 4 assessment days to better determine the impact of work on fatigue diurnal patterns within-subjects and, indeed, fatigue recall. There were no differences in employment profiles between groups, and assessment days were restricted to weekdays, both of which should have minimised this limitation.

Some other daily factors that may affect MS fatigue were not measured. For instance, MS fatigue has been described as sensitive to heat exposure, sunlight, and humidity (Stuifbergen & Rogers, 1997). The impact of heat and climate on fatigue could not be determined in the present study. A recent

study reported 53.4% of 88 pwMS were classified as experiencing heat sensitive fatigue by self-report (Bol et al., 2012). Participants were classified as "heat sensitive" if they scored greater than 5 on two 1-7 VAS items: "Heat brings on my fatigue", "Cool temperatures lessen my fatigue". Despite this, Bol et al. (2012) could find no relationship between individuals' local climatological data obtained via an online resource (mean and maximum daily temperature, estimates of UV exposure) and mental or physical fatigue. However, patients may perceive the heat sensitive nature of fatigue by fluctuations within seasons or within days, rather than by general climates.

An iron-deficiency anaemia diagnosis led to participant exclusion, but it was not possible to measure serum levels of iron as a covariate; doing so would have involved invasive methods potentially invoking stressful experience and eliminating a lot of the benefits already described in carrying out a study in everyday life. Iron supplements have been shown to improve unexplained fatigue even in non-anaemic iron-deficient women in previous research (Vaucher, Druais, Waldvogel, & Favrat, 2012; Verdon et al., 2003). No participant was taking iron supplements at the time of the EMA study, and the clinics in Southampton and London confirmed that iron deficiency was tested routinely after every clinic.

Although compliance with the EMA protocol (91%) was considered excellent, there were still some missing assessments possibly introducing some bias. While there was no reason to suggest missing events were not missing at random (see section 2.2.6.4.2 of Chapter 2), it was plausible that some missing events were due to excessive fatigue at the time of assessment. Being prompted by auditory alarm was considered to minimise this potential bias, but cannot be completely discounted as the information required was missing as a result. It would be useful to be able to quantify the reasons for missing assessments (Gil et al., 2003; Porter, Gil, Carson, Anthony, & Ready, 2000) but this would rely on retrospective evaluations which could themselves be influenced by recall bias and social desirability. A retrospective evaluation of reasons for missing assessments was not carried out here.

Fatigue was not measured first thing in the morning. There has been evidence that waking levels of fatigue may be similar between a clinical group (cancer) and control group, only for fatigue to be elevated later in the day for those in the clinical group (Curran et al., 2004). Fatigue was not measured at

awakening in the present study so as not to confound fatigue assessments with sleep-wake mechanisms. Secondly, it was preferable to measure MF repeatedly within the same variable-occasion A1-A6 design rather than include a single additional morning assessment within the fixed morning S1-S3 design. This retained consistent expectancies and consistent anticipatory effects for each assessment.

A possible ceiling effect in the MF measure which may have produced the quadratic trajectory in the RRMS group was not considered a significant limitation; the maximum score of 10 on this item was used infrequently (< 3% of MF assessments in the RRMS group). Therefore it is unlikely or, at least, infrequently encountered, that a participant had reached a measurement ceiling when providing an MF rating. This position is further substantiated by the presence of a decelerating diurnal fatigue pattern in the RRMS-nf group where overall fatigue levels were substantially lower than the RRMS-f group.

Examination of peak and end biases in fatigue recall reports was limited to daily recalls. Future research should examine longer recall windows of a week or even a month to investigate the accuracy of these reports and build on the results presented here. In a study of pain recall in rheumatology patients, pain recall was shown to be less accurate with extended recall periods up to 7 days, only for 28-day recall to be more accurate than 7-day recall (Broderick et al., 2008). Determining how fatigue reports are provided in MS has important implications for both research and clinical practice.

Finally, it is likely that some peaks in fatigue severity were missed due to the infrequency of reports during the day. “End” ratings of fatigue were also provided at the quasi-random event prior to the EOD event at 2100h. It is possible that the time elapsed between the MF assessment provided within the A6 event and the EOD assessment may have masked some of the end heuristic in recall provision.

3.4.4 Conclusion

This chapter presented evidence that fatigue severity and impact in pwRRMS is greater than in healthy individuals at different levels of measurement (momentary, daily recall, monthly recall). The diurnal fatigue pattern in everyday life follows an increasing trajectory in RRMS which

decelerates with time and peaks in the late afternoon/early evening. This decelerating daily fatigue trajectory was not present for healthy participants. Daily contextual factors (stress and mood) were associated with fatigue in both groups, but only stressors defined as hindrances to goal achievement predicted greater fatigue. Stressors that may facilitate goal achievement were not associated with fatigue.

Although discrepancy between daily recalled fatigue and aggregated daily momentary assessments was relatively small in both groups, there was evidence of a small but significant use of peak and end memory heuristics in daily recalled fatigue in RRMS not present in healthy individuals. Future research is required to investigate their influence over larger recall periods in MS.

Chapter 4: Cortisol secretory activity in relapsing-remitting MS in everyday life

MS is a disease characterised by autoimmune-mediated inflammatory demyelination and neurodegeneration. Cortisol is an endogenous anti-inflammatory and there has been evidence that stress and dysfunction of the HPA axis is highly relevant to the control and course of MS disease processes (Heesen, Gold, et al., 2007; Heesen, Mohr, et al., 2007; Mohr et al., 2004). This chapter begins by reviewing studies of cortisol secretory activity (CSA) in MS, drawing on literature using different methodologies. The chapter goes on to describe an analysis exploring CSA in everyday life in RRMS and healthy control groups, including the cortisol awakening response (CAR), diurnal cortisol slope (DCS) and cortisol reactivity to daily stressors.

4.1 Introduction

4.1.1 Pharmacological challenge tests

Until very recently, research of HPA axis function in MS predominantly used the dexamethasone suppression test (DST) (Liddle, 1960) or the combined dexamethasone-corticotrophin releasing hormone (Dex/CRH) test (Heuser et al., 1994) to examine HPA axis regulation and sensitivity of glucocorticoid receptors. In these tests, cortisol non-suppression following dexamethasone and hyper-responses following CRH provocation indicate dysfunction (see Appendix B for standard protocols and rationale).

Studies using the standard DST protocol have found moderate prevalence (33-62%) of HPA axis dysfunction in MS defined as non-suppression of dexamethasone (Heesen, Gold, et al., 2002; Reder et al., 1987; Ysrraelit et al., 2008); the two studies reporting low rates of non-suppression (0-12%) in MS delayed morning samples by 30-60 minutes (Then-Bergh et al., 1999; Wei & Lightman, 1997). However, the determination of cortisol at a single time-point in the DST is a particular shortcoming given cortisol is a fluctuating, pulsatile hormone. Therefore, research has tended to move away from the DST to instead examine feedback mechanisms using the more sensitive Dex/CRH test,

computing markers such as $\Delta max_{cortisol}$ (subtracting baseline from peak cortisol level) to assess HPA axis hyper-response to CRH under suppressed conditions.

Then-Bergh et al. (1999) conducted the Dex/CRH test with 60 pwMS (38 RRMS, 16 SPMS, 6 PPMS) and 29 age- and sex-matched healthy controls. However, while the small PPMS subgroup presented large hyper-responses ($\Delta max_{cortisol}$) compared to the control group, d = 2.39, 95% CI [1.30, 3.37], cortisol responses in the RRMS group, d = 0.19, 95% CI [-0.30, 0.67], and SPMS group, d = 0.38, 95% CI [-0.24, 1.00], were similar to the control group. Heesen, Gold et al. (2002) also found no statistically significant hyper-responses to the Dex/CRH test in either PPMS (n = 13) or RRMS (n = 8) compared to controls (n = 11). An additional study by Schumann et al. (2002) carried out the Dex/CRH test in an MS group (n = 55; all MS types, ratio unclear) and healthy control group (n = 29), but found no difference in $\Delta max_{cortisol}$, d = 0.17, 95% CI [-0.28, 0.62].

While both Then-Bergh et al. (1999) and Heesen, Gold, et al. (2002) found Dex/CRH test outcomes were similar in RRMS and controls, two other studies reported cortisol hyper-responses to the Dex/CRH test in RRMS (Fassbender et al., 1998; Grasser et al., 1996). Fassbender et al. (1998) found greater $\Delta max_{cortisol}$ in an RRMS group (n = 23) than a healthy control group (n = 17), d = 0.93, 95% CI [0.25, 1.57], and Grasser et al. (1996) reported a similar difference between an RRMS group (n = 19) and control group (n = 19), d = 0.81, 95% CI [0.13, 1.45]. Identical protocols were carried out in all Dex/CRH test studies described and, in all studies, participants were at least 4 weeks from any corticosteroid treatment. RRMS group participants were typically in relapse-phase (heightened inflammation) in all studies.

In a meta-analysis examining age as a moderator of cortisol responses to pharmacological challenge tests, older individuals (meta-analysis mean age = 69.1 ±6.5 years) demonstrated larger responses than younger people (mean age = 27.8 ±5.1 years) with a mean effect size of d = 0.42 (Otte et al., 2005). It is possible that the effect size reported by Fassbender et al. (1998) was inflated due to an older RRMS group; although the authors reported no statistically significant between-group age-differences, the reported age ranges indicated otherwise: 35-62 years for the RRMS group; 24-39 for the control group. There is also evidence that MS disease activity may be relevant to Dex/CRH test results, although the direction of the effect is unclear:

Fassbender et al. (1998) found positive associations (r = .63) between Dex/CRH test hyper-responses and white blood cell count in cerebrospinal fluid (marker of CNS inflammation); but Schumann et al. (2002) reported a negative association (r = -.44) between hyper-responses and number of Gd+ lesions on MRI. Disease activity may account for some of the differences in results between studies, but the nature of its effect is unclear.

Depression among pwMS may also contribute to HPA axis dysregulation. Depression has been frequently associated with HPA axis hyperactivity in the past (Knorr et al., 2010; Pariante & Lightman, 2008; Stetler & Miller, 2011), and lifetime prevalence is around 50% in MS (Goldman Consensus Group, 2005). Fassbender et al. (1998) included four RRMS participants who fulfilled DSM-III criteria for major depressive disorder (MDD), and reported a mean score of 9.7 on the Hamilton Rating Scale for Depression (Hamilton, 1960) in the RRMS group, reflecting mild-moderate symptoms. The two studies showing little HPA axis dysfunction in RRMS reported low depressive symptoms in their samples: Then-Bergh et al. (1999) reported the Hamilton Depression scale mean was 2.0, while Heesen, Gold, et al. (2002) stated "depression scores [HADS-D] in our sample were in the lower range..." (p.513).

Neither Then-Bergh et al. (1999) nor Heesen, Gold, et al. (2002) could identify statistically significant correlations between endocrine markers and depression scores in MS (r < .20), arguing results suggested little support for depression determining HPA axis dysregulation in MS. However, Fassbender et al. (1998) found a positive association between $\Delta max_{cortisol}$ and Hamilton Depression scores (r = .56), so a role for MDD and/or depressive symptoms in HPA axis hyperactivity in MS cannot be ruled out. It is possible the studies showing small associations reflected a floor effect due to a low range (variability of scores). There was no depression data reported by Grasser et al. (1996) or Schumann et al. (2002). Recently, Gold et al. (2011) published a study including the DST carried out with pwRRMS with (n = 10) and without (n = 34) comorbid MDD. There were no differences between groups for the DST; however, differences in basal salivary cortisol measures were found, which will be discussed in detail in the following section.

Evidence is mixed regarding HPA axis hyper-responsivity as determined by pharmacological challenge tests in MS and in RRMS specifically. Where HPA axis dysfunction has been reported in MS, the size of hyper-response may

depend on the extent of disease activity or the presence of depressive comorbidity. However, currently the published literature is equivocal on these points.

4.1.2 Cortisol reactivity to psychosocial stressors in MS

A meta-analysis of physiological responses to experimental psychosocial stressors demonstrated substantial heterogeneity in HPA axis responses, concluding that tasks containing socio-evaluative threat (possibility of negative assessment of task performance) and uncontrollability reliably elicited the largest cortisol and ACTH responses (Dickerson & Kemeny, 2004). Only two studies have examined the HPA axis response to psychosocial stress in MS (Ackerman et al., 1996; Heesen, Schulz, et al., 2002), despite accumulating evidence associating stressful life events with increased risk of symptom relapse (exacerbation) in RRMS (Artemiadis et al., 2011; Mohr et al., 2004).

Ackerman et al. (1996) included an MS group (n = 25; remission-phase RRMS and SPMS) and an age- and sex-matched healthy control group (n = 25). MS participants had not received corticosteroids or antidepressants for 2 months, had no history of DMT-use, and the final sample of 25 was derived by excluding individuals with low scores (<5th percentile) on the Screening Examination for Cognitive Impairment (Beatty et al., 1995). The experimentally-induced psychosocial stressor was a 3 minute "videotaped" oral defence against the hypothetical situation that they had found a wallet in a store but had been accused of stealing the wallet by security. Participants were given 2 minutes to prepare their speech, and their performance was "rated" by an observer to increase situational anxiety (socio-evaluative threat). The study was carried out in the laboratory between 0700h and 1000h for all participants. Blood samples were drawn via a catheter from the forearm at baseline (t0), 5 (t1), 20 (t2), and 60 minutes (t3) after stressor onset. Immediately after the speech (t1), participants rated eight items (stressful, involving, frustrating, challenging, engaging, irritating, difficult, interesting) on a 4-point Likert scale to assess subjective response to the stressor. At t2, participants were fully debriefed about the nature of the study and asked to sit and relax until the t3 sample was taken.

Cortisol at t0 was, on average (for both groups), 17.0 ±5.0 nmol/L, followed by 16.4 ±5.0 nmol/L at t1, 17.9 ±5.0 nmol/L at t2, and 13.3 ±4.0

nmol/L at t3. The report notes small statistically significant differences between cortisol levels at t2 compared to t1 (but not t0), and lower cortisol levels at t3 compared to all other samples. Both groups showed equal activation scores, and an increased mood disturbance (measured by the Profile of Mood States; McNair, Lorr, & Dropleman, 1971) following the stressor. Although the stressor did not incorporate uncontrollability, it did provide socio-evaluative threat. Therefore, the stressor would be expected to elicit a cortisol response after baseline, but it did not. The most notable methodological problem with the study is the time of day of participation. It would have been more appropriate to control for the cortisol circadian rhythm by consistently performing the protocol in the afternoon when cortisol levels are more stable; the early morning is affected by the natural decrease in cortisol levels following the CAR.

The study by Heesen, Schulz, et al. (2002) had the same flaw, conducting all acute psychosocial stressor tasks at 0830h. Thirty-five pwRRMS were randomly assigned to either a "stress" (n = 20) or a "no-stress" (n = 15) group. The "stress" group, and a group of healthy controls (n = 15), were required to perform a 45-minute multifaceted stressor task including an arithmetic task, Stroop task, and video-recorded public-speaking task. Exact details of the tasks were not reported, but the overall task was substantiated by elevated heart rate immediately after the task compared to baseline. The "no-stress" group protocol was not reported. Plasma cortisol levels were measured before (t0), immediately after (t1), and 1 hour after the tasks (t2). As Dickerson and Kemeny (2004) had detailed that tasks involving public speaking/cognitive task combinations were associated with the greatest cortisol responses across studies, Heesen, Schulz, et al. (2002) expected similar effects. However, results showed reductions in cortisol levels in both MS (t0, 251 ±85 pg/ml; t1, 201 ±89 pg/ml; t2, 157 ±69 pg/ml) and control (t0, 208 ±83 pg/ml; t1, 202 ±52 pg/ml; t2, 144 ±48 pg/ml) "stress" groups across time, and significant reductions from baseline cortisol in the MS "no-stress" group (t0, 306 ±181 pg/ml; t1, 224 ±152 pg/ml; t2, 227 ±142 pg/ml).

The only two studies published examining cortisol reactivity to stress in MS both have the same methodological flaw. Dickerson and Kemeny (2004) found effect sizes were diminished in morning stressor tasks (d = 0.14) compared to the afternoon (d = 0.46). Conducting the study in the morning

was further compounded by neither reporting nor controlling for time of awakening. In addition, neither study reported cortisol levels prior to the onset of the stressor when it is now well-established that anticipation of a stressor can elicit a cortisol response (Balodis, Wynne-Edwards, & Olmstead, 2010; Engert et al., 2013; Kirschbaum, Wüst, & Hellhammer, 1992; Mason et al., 1973).

4.1.2.1 Daily hassles

Reactivity of the HPA axis in MS to experimentally-induced acute psychosocial stressors has not yet been adequately addressed, and reactivity to everyday life stressors in MS has yet to be examined. Daily hassles and stressful events have previously been shown to be sufficient to elicit a cortisol response in the general population (for example, Jacobs et al., 2007; Schlotz et al., 2006; J. Smyth et al., 1998; van Eck, Berkhof, Nicolson, & Sulon, 1996) and in clinical populations (for example, Peeters et al., 2003). Investigations of cortisol reactivity to stress could help to elucidate mechanisms underlying the apparent link between stress and MS disease course (Mohr et al., 2004; Mohr & Pelletier, 2006). Inadequate responses could thwart the achievement of homeostasis as deficient endogenous cortisol levels is associated with inflammation and other physical disorders (Heim et al., 2000).

4.1.3 Basal cortisol secretory activity in MS

Several studies have examined unstimulated CSA in MS, using methods including 24h urinary cortisol summary measures and momentary serum or saliva assessments. When serum and saliva samples are provided in repeated measures designs, they permit the measurement of different facets of cortisol's circadian rhythm.

4.1.3.1 24h urinary cortisol and serum cortisol

Michelson et al. (1994) measured 24h urinary cortisol in 13 pwMS (10 RRMS; 3 progressive) and 12 healthy individuals, over 2 or 3 days. One participant in the MS group (MS-type not reported) was removed due to sample loss. Mean 24h urinary cortisol was significantly higher in pwMS than controls, d = 1.05, 95% CI [0.16, 1.86]. Evening serum cortisol and ACTH levels were also observed by aggregating samples taken at 15-minute intervals between 1800h and 2000h. Evening cortisol levels were significantly elevated in pwMS

compared to controls, *d* = 1.50, 95% CI [0.55, 2.35]; however, ACTH levels were slightly lower in pwMS, *d* = -0.50, 95% CI [-1.29, 0.33], indicating a dissociation between cortisol and its ACTH precursor in the evening, and suggesting HPA axis dysfunction at the adrenal cortex in MS.

Ysrraelit et al. (2008) reported a more comprehensive study with 173 pwMS and 60 healthy individuals, examining HPA axis activity over a single day with (1) 24h urinary cortisol; and (2) a single serum sample for cortisol and ACTH in the morning (0800h). The MS sample was heterogeneous and divided into four groups: PPMS (*n* = 40); SPMS (*n* = 41); remission-phase RRMS (*n* = 58); and relapse-phase RRMS (within 5 days of an attack, *n* = 34). The healthy control group was divided into two equal-sized groups, with one group recruited matching the progressive MS groups for age and sex (older and more equal gender ratio), and the other matching the RRMS groups. Results showed the remission-phase RRMS group, *M* = 294.1 ±70.0 µg/24h, *d* = 0.68, 95% CI [0.23, 1.13], and relapse-phase RRMS group, *M* = 441.5 ±67.9 µg/24h, *d* = 2.55, 95% CI [1.86, 3.18)], presented higher 24h urinary cortisol levels than their respective control group, *M* = 241.6 ±88.8 µg/24h. The PPMS group, *M* = 348.3 ±115.6 µg/24h, *d* = 2.03, 95% CI [1.42, 2.58], and SPMS group *M* = 271.8 ±65.7 µg/24h, *d* = 1.92, 95% CI [1.33, 2.46], also secreted more cortisol over the 24h period than their corresponding control group (*M* = 162.6 ±41.8 µg/24h).

Serum cortisol and ACTH at 0800h were both elevated in all MS groups, which did not support the ACTH-cortisol dissociation observed in the evening by Michelson et al. (1994). However, 0800h serum measures did not account for the time of awakening and associated CAR, making them potentially unreliable. The article refers to samples being made between 0800h and 1000h, yet refers only to "0800h" samples in the results.

Standardised effect sizes were largest in those MS groups characterised by heightened disease activity: relapse-phase of RRMS and progressive MS-types. This observation likely reflects a regulatory response by the HPA axis to increased inflammation, although cortisol output was still greater than controls for RRMS participants in remission. An observation overlooked by the authors was the large difference in 24h urinary cortisol between the respective control groups, *d* = 1.14, 95% CI [0.58, 1.67]. The two control groups differed in mean age (33.9 versus 49.0) and female to male ratio (23:7 versus 16:14), but the

higher 24h cortisol was in the younger group. This observation is surprising as HPA axis activity is usually elevated in older individuals and is typically greater in women than men (Larsson, Gullberg, Rastam, & Lindblad, 2009; Otte et al., 2005).

Studies examining basal CSA in MS using urinary and/or serum measures have consistently presented evidence of HPA axis hyperactivity in RRMS, and even more so in progressive MS-types. However, the methods and sampling frequency employed lacked the sensitivity to observe variability in a pulsatile hormone secreted in a circadian rhythm, and serum cortisol sampling provided in unfamiliar contexts with invasive procedures (venepuncture) may not be generalisable to everyday life.

4.1.3.2 Salivary cortisol

Since 2010, there have been four published studies in RRMS of salivary cortisol in everyday life from two research groups (Gold et al., 2010; Gold et al., 2011; Kern et al., 2013; Kern et al., 2011). In the first study of its type in MS, Gold et al. (2010) used a case-control design requesting participants provide three salivary cortisol assessments per day (awakening; 1600h; 2100h) over 2 consecutive days. Cortisol levels were aggregated across days for each time-point for group comparisons. Although cortisol levels were no different in the RRMS group (n = 29) than the healthy control group (n = 20) upon awakening and at 1600h, there was a small, statistically significant elevation in 2100h cortisol levels in the RRMS group (mean and *SD* not reported). The DCS, computed as the linear regression slope of log-transformed values, appeared somewhat flatter in the RRMS group than the control group but did not quite reach statistical significance (p = .08).

In an exploratory analysis, Gold et al. (2010) tested the hypothesis that depressive symptoms moderate HPA axis hyperactivity in RRMS. The RRMS group was divided into those with (n = 8) and without depression (n = 21) based on a ≥14 cut-off on BDI-II; scores of 14 and above on this scale are indicative of at least mild depression (Beck et al., 1996). Both flatter DCS and higher 2100h cortisol was evident in the RRMS BDI-high group compared to the RRMS BDI-low group and healthy control group. All differences were to statistical significance, but data was not reported to compute effect sizes.

There was no difference between the BDI-low group and the healthy control group on any cortisol outcome.

Finally, Gold et al. (2010) examined the importance of antidepressant medication to cortisol levels in RRMS by comparing cortisol outcomes in the control group to those in the RRMS BDI-low group taking antidepressants (n = 5; indicative of "well-controlled depression") and the RRMS BDI-high group not taking antidepressants (n = 6; "unmedicated depression"). Despite very small group sizes, a trend towards a flatter DCS was reported in the "unmedicated depression" group compared to the healthy control group (p = .07). Those with "well-controlled depression" were described as having DCS similar to healthy controls and others in the RRMS BDI-low group. It would appear that the main effect was largely an artefact of how the researchers allocated individuals taking antidepressants, as these individuals had the steepest DCS (steeper than controls, on average). It would seem counterintuitive to include more individuals taking antidepressants in the "no depression" group (n = 5) than the "depression" group (n = 2); although it is possible antidepressants were taken for neuropathic pain in MS (Dworkin et al., 2007). If researchers had allocated groups based on clinical diagnosis of MDD, more individuals taking antidepressants would likely have been in the "depression" subgroup, and the conclusion associating flatter DCS with depression in MS may not have been reached.

Gold et al. (2010) was the first study to examine the repeated measurement of CSA in everyday life in MS. The study recruited a relatively homogeneous MS sample (remission-phase RRMS) which was particularly important given the small sample size and differences in pathophysiology of RRMS and progressive MS types demonstrated elsewhere (Ysrraelit et al., 2008). However, eligibility criteria did not exclude physical or psychiatric comorbidities which, given the high levels of comorbidity in MS (Marrie et al., 2008), means findings may not necessarily be attributable to RRMS itself. With the very small group sizes in some analyses, findings should be interpreted with caution.

The reliance on the BDI-II cut-off for depressive comorbidity instead of a clinical diagnosis of MDD was a major criticism of the Gold et al. (2010) study, as well as the lack of tight exclusion criteria to ensure findings were attributable to MS and/or depressive symptoms. Gold et al. (2011) later

reported a study with 44 pwRRMS with (n = 10) and without (n = 34) a clinical diagnosis of MDD. This study had no healthy control group. MDD was identified by a clinical psychologist using DSM-IV criteria, no participant was taking antidepressants at the time of the study, and the study excluded individuals with a history of endocrine abnormalities or other psychiatric disorders. All participants were female. Participants were asked to collect saliva samples over 2 consecutive days incorporating measurement of the CAR with assessments at awakening (T0), 15 minutes after awakening (T15), T30, T45, and T60; and circadian profile cortisol output and DCS by T0 cortisol and further samples at 1100h, 1500h, 2000h, and 2300h. Results showed the RRMS+MDD group had a larger average circadian profile AUCg than the RRMS group, d = 0.74, 95% CI [0.00, 1.45], as well as a flatter DCS, seemingly as a result of a statistically significant elevation in evening cortisol level in the RRMS+MDD group (data not reported). The difference in DCS was retained even after controlling for neurological disability (EDSS); whether this held true for the AUCg was not stated by the authors. The CAR AUCg was similar between groups. Therefore, both studies by Gold and colleagues suggested a flatter DCS is characteristic of depression within MS, rather than MS itself.

The comorbid depression hypothesis for HPA axis hyperactivity in MS was further supported in a study by Kern et al. (2011) with 32 pwRRMS and 16 healthy controls. The RRMS group was split into two groups by median split on the BDI-II (median = 7). It was unclear whether any participants were taking antidepressants. Saliva samples were collected over 2 (not necessarily consecutive) days within 1 week at awakening (T0), T30, T45, and T60 to measure the CAR, and also at 1500h and 2200h. In non-split group comparisons, CAR AUCg was significantly elevated in pwRRMS compared to controls, d = 0.76, 95% CI [0.13, 1.36], and statistically significant differences between groups for all morning assessments except T0, which may itself also have been statistically significant with greater power (T0, d = 0.33; T30, d = 0.59; T45, d = 0.91; T60, d = 0.81). Cortisol levels were similar between groups at 1500h and, in contrast to the findings of the Gold group, the study did not find elevated cortisol levels in RRMS in the evening (2200h) (M = 2.49 ±1.39 nmol/L) compared to a healthy control group (M = 2.51 ±2.04 nmol/L).

Compared to the healthy control group, CAR AUCg was elevated in the RRMS BDI-high group, d = 1.01, 95% CI [0.25, 1.72], but not the RRMS BDI-low

group, *d* = 0.55, 95% CI [-0.17, 1.24]. It appeared from the study report that a lack of statistical power disguised a probable difference between the RRMS BDI-low group and control group. The 95% CI for this group comparison was wide (-50.5 nmol/L to 436.2 nmol/L), suggesting data from more participants would be needed before being confident that CAR differences between pwRRMS and healthy controls could be solely attributed to depression. Further, scoring above 7 on the BDI-II is arbitrary and means little clinically (a score above 14 is required to indicate mild depression), and using a median split results in a loss of statistical power (MacCallum, Zhang, Preacher, & Rucker, 2002; Naggara et al., 2011). The number of individuals in the BDI-high group with scores >14 is unknown, but the mean BDI-II score for the BDI-high group was 12.47 ±4.11. Cortisol at 1500h and 2200h was similar between RRMS BDI-high, RRMS BDI-low, and control groups.

Very recently, Kern et al. (2013) published another study with a larger sample size, recruiting 55 pwRRMS and 34 healthy control participants, as well as 22 pwSPMS. The 2-day sampling protocol matched that used in the research group's earlier study (Kern et al., 2011), but depression was measured by the Centre for Epidemiological Studies Depression (CES-D) Scale (Radloff, 1991) and median split was not used. The study found support for a difference in CAR AUCg between RRMS and healthy controls, *d* = 0.57, 95% CI [0.13, 1.00], but there were no differences in CAR AUCg between RRMS and SPMS, *d* = -0.44, 95% CI [-0.94, 0.06], nor between SPMS and healthy controls, *d* = 0.15, 95% CI [-0.39, 0.69].

Although both MS groups had significantly higher CES-D scores than the control group, depressive symptoms were not associated with CAR AUCg in either RRMS (*r* = .02, *p* > .05) or SPMS (*r* = -.13, *p* > .05). The authors presented only T-scores for CES-D scores in RRMS (T = 53; 0.3 *SD* above population mean), so the absolute level of depressive symptoms in the sample was unclear, but the authors state that overall CES-D scores were relatively low in the MS groups and not clinically significant. These results suggest differences in basal CSA in RRMS exist in the absence of clinically relevant depression.

Kern et al. (2013) were also interested in the effects of DMTs, chronic stress (TICS), and neurological disability (EDSS) on CSA. People with RRMS who had used DMTs at any stage did not differ from "treatment naïve" patients for

CAR AUCg; however, only DMT-treated pwRRMS differed from healthy controls to statistical significance (RRMS-treated vs controls, d = 0.62; RRMS-naïve vs controls, d = 0.49). CAR AUCg was not associated with neurological disability (r = .09, p > .05) or chronic stress (r = -.02, p > .05) in RRMS. However, CAR AUCg was associated with neurological disability (r = .46, p < .05) and chronic stress (r = -.48, p < .05) in SPMS, suggesting chronic stress is not relevant to the CAR in RRMS, but may be more important in progressive stages of the disease. The negative association with chronic stress (TICS score) in SPMS and the lack of a relationship in RRMS is surprising given TICS scale scores are usually positively associated with the CAR AUCg in healthy populations (for example, M. Pruessner, Hellhammer, Pruessner, & Lupien, 2003; Schlotz et al., 2004) and may therefore represent a dysfunction of the HPA axis stress response system in MS. Examining whether anticipated positive associations between TICS scores and CAR AUCg were present in the healthy control group would have been informative, but data was not reported.

Finally, a 9-month follow-up conducted with the RRMS group only was also used to define individuals as "symptom-stable" (defined as EDSS progression from baseline to 9-month ≤0; n = 17) or "symptom-progression" (EDSS progression ≥0.5; n = 9). Group comparisons revealed the "symptom-stable" group did not differ from healthy controls for CAR AUCg at baseline, but the "symptom-progression" group had larger CAR AUCg (d = 0.92). Although this suggests those with heightened disease activity had stronger CARs, it should be noted that the follow-up attrition rate (53%) was quite high and those with worsening neurological symptoms may have been more likely to drop out.

Overall, these four studies have approached many relevant research questions in an ecologically valid manner. All studies were strict in terms of recruiting homogeneous RRMS groups; all at least 4 weeks from steroid treatment and acute relapse. However, all three of the case-control studies (Gold et al., 2010; Kern et al., 2013; Kern et al., 2011) did not explicitly report exclusion of physical or psychiatric comorbidities, making it impossible to be sure results were not confounded by effects on HPA axis activity by other medical conditions or treatments. There were several other common limitations. Firstly, although each study assessed cortisol upon awakening, no study monitored how accurate participants were in providing timely samples,

and do not describe means of maximising sampling protocol compliance. Given the importance of measuring the awakening sample immediately is now well-established (Griefahn & Robens, 2011; Okun et al., 2010; N. Smyth et al., 2013), an attempt to ensure compliance in everyday life studies of CSA is vital. Throughout the rest of the day, non-adherence can also introduce biases; for example, some participants may delay providing a sample if they are experiencing high levels of stress or symptoms.

The studies also sampled over a maximum of 2 days, and aggregated cortisol data for each time-point across days, removing day-to-day variance and preventing control for daily confounding factors such as sleep quality. The fixed-event design also failed to provide proper account of the whole day, where a random-event design would have led to samples spread throughout the day in a more representative manner (Schlotz, 2011). Of course, a random-event design would only be possible with electronic prompting of samples, which none of these studies incorporated. A study in MS taking steps towards maximising participant compliance with protocols, thus minimising the biases associated with taking studies into the "real-world", has yet to be published.

4.1.4 Aims

The primary aim of this investigation was to comprehensively explore salivary cortisol outcomes in everyday life, including the CAR and daytime CSA using a prospective case-control design. The following hypotheses were tested: (1) The CAR is elevated in RRMS than healthy individuals; (2) CAR elevation in RRMS is moderated by depressive symptoms, such that CARs are larger in those with high depressive symptoms scores; (3) The DCS is flatter in RRMS than healthy individuals as a result of elevated evening cortisol; (4) Cortisol reactivity to stressors in everyday life is attenuated in RRMS compared to controls.

4.2 Method

General methods were presented in section 2.2.2 of Chapter 2. Features of the EMA protocol relevant to this analysis are presented in brief here for convenience, followed by a detailed overview of the statistical procedures carried out to test hypotheses.

4.2.1 Participants

The study recruited an RRMS group (*n* = 38) and an age- and sex-matched healthy control group (*n* = 38). Full descriptive statistics for the study sample were provided in Table 2 of Chapter 2 (p.51).

All RRMS participants had a neurologist-confirmed RRMS diagnosis, were at least 3 months from a clinical relapse and/or corticosteroid treatment, had no other physical or psychiatric comorbidities, and were not taking antidepressants. Full details of eligibility criteria were provided in section 2.2.1 of Chapter 2.

4.2.2 Design

Participants reported to the laboratory to provide informed consent, complete baseline questionnaires, and undertake thorough training in using the handheld (Hewlett Packard iPAQ 111 Classic Handheld*, Bracknell, UK*) and providing saliva samples using Cortisol Salivettes (Sarstedt*, Leicester, UK*). EMA took place over 4 consecutive weekdays to control for weekday-weekend effects on CSA (Kunz-Ebrecht et al., 2004; Schlotz et al., 2004; Thorn, Hucklebridge, Evans, & Clow, 2006). EMA was prompted by auditory alarm and instructed via touchscreen display on the handheld. The two facets of the EMA design relevant to this chapter were the fixed-occasion event-based morning design (S1–S3 events) and the variable-occasion quasi-random design (A1–A6 events, 1000h–2000h).

4.2.2.1 Cortisol assessment design

The CAR was measured using saliva samples obtained at events S1 to S3: awakening (S1/T0); 30 minutes after awakening (S2/T30); and 15 minutes later (S2/T45). The S1 event was initiated in one of three ways: (1) pre-set auditory alarm; (2) automatic auditory alarm at 0830h if not already awake; or (3) manual initiation upon awakening (pressing "Start" on the handheld screen). Until completion of the S3 event, participants were instructed to refrain from eating, brushing teeth, engaging in intensive physical activity, and smoking, and to only drink water. This is a standard set of behavioural restrictions for the reliable measurement of the CAR (Chida & Steptoe, 2009). Other than these restrictions, participants were free to engage in usual morning routines.

The DCS was measured by saliva samples obtained at events A1 to A6, which permitted collection of a representative sample of daily CSA (between 1000h and 2000h) and the examination of within-subjects contextual effects (Kudielka et al., 2012; Schlotz, 2011). Participants were not expected to refrain from eating or drinking prior to sampling; however, these behaviours were monitored via momentary self-report and accounted for in statistical analyses.

4.2.3 Measures

Section 2.2.3 of Chapter 2 provided justification for and critical examination of all baseline and momentary assessment measures used in this investigation. Those of relevance are detailed in brief below.

4.2.3.1 Baseline measures

Baseline measures of relevance to this chapter were the self-administered Expanded Disability Status Scale (EDSS; Bowen et al., 2001), the Chronic Stress Screening Scale (CSSS; Schulz et al., 2004), and the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D; Zigmond & Snaith, 1983). Socio-demographic and other personal measures (age, gender, ethnicity, socioeconomic status, menstrual phase, oral contraceptive use, DMT) were assessed for examination as potential covariates. Details of these scales, their validities and justifications were discussed in section 2.2.3.1 of Chapter 2.

4.2.3.2 EMA measures

4.2.3.2.1 Cortisol

Cortisol was assessed using saliva samples obtained using Cortisol Salivettes, which is easily self-administered and frequently used in everyday life studies (Kirschbaum & Hellhammer, 1989; Kudielka et al., 2012). To maximise compliance, handhelds briefly presented a 3-digit code upon each event alarm which participants recorded on the label of the salivette used. This method has been successfully implemented into previous cortisol studies (>90% adherence) (Powell & Schlotz, 2012; Stetler & Miller, 2011).

Participants stored used salivettes in a refrigerator at the end of each day. Upon completion of sampling, salivettes were returned to the researcher and immediately stored in a freezer at -20°C until data collection was completed. Salivary cortisol samples are stable at room temperature for up to 4 weeks

(Kirschbaum & Hellhammer, 1989) and at -20°C for periods of up to 12 months (Garde & Hansen, 2005).

Three markers of the CAR were used: (1) AUCg; (2) AUCi; and (3) S1 cortisol level. AUCg and AUCi were computed using formulae presented by J. C. Pruessner et al. (2003) detailed in Appendix G. CAR AUCg is a measure of estimated total cortisol output, CAR AUCi measures the change in cortisol output post-awakening, and S1 cortisol is an indicator of cortisol level at the end of the pre-awakening rise (Clow et al., 2010; Fekedulegn et al., 2007; D. H. Hellhammer, Wüst, & Kudielka, 2009). The pre-awakening period is thought to be characterised by reduced adrenal sensitivity to ACTH, mediated by the suprachiasmatic nucleus and potentially modulated by the hypothalamus (Clow et al., 2010; D. H. Hellhammer et al., 2009); ACTH has been shown to rise more quickly that cortisol in the pre-awakening period in humans (Born, Hansen, Marshall, Molle, & Fehm, 1999). Clow et al. (2010) suggested post-awakening cortisol is associated with heightened rather than reduced adrenal-sensitivity to ACTH, mediated by suprachiasmatic nucleus light-sensitivity. This was supported by dawn simulation waking (gradually increasing light levels before waking up) leading to increased post-awakening cortisol output (but not increased S1) compared with waking in darkness (Thorn, Hucklebridge, Esgate, Evans, & Clow, 2004). The DCS was computed as the linear slope estimated by the multilevel model of log-transformed cortisol samples collected at events A1-A6. Computing AUC was not appropriate due to the variable-occasion assessment design.

4.2.3.2.2 Self-report

Self-report items were presented via the handheld. Sleep quality was measured at event S1 (awakening) with the anchor points “Very bad” (0) and “Very good” (10). At S1, participants also reported the time of going to sleep, with length of sleep derived as the time elapsed between this self-report and the S1 timestamp. At each quasi-random event (A1–A6), behavioural activities (physical activity, drinking coffee, eating a meal, and smoking) were assessed by check-box response (Yes/No).

Momentary stressors and mood were assessed at each quasi-random assessment (A1-A6) event by VAS with anchor points “Disagree” (0) and “Agree” (10). Momentary stressors were measured by eight items, preceded by

the words "Since the last event…" An exploratory factor analysis (see section 2.2.3.2.3.2 of Chapter 2) yielded two subscales: *Stressor Hindrance* (threatening constraints that frustrate needs and goal achievement) and *Stressor Challenge* (workload, both physical and social, which may or may not contribute towards goal achievement). Participants could respond "Not applicable" to stressor items, resulting in a missing value. Subscale scores were computed as the mean of constituent items. Morning stress covering the period "Since waking up today…" was assessed at the S3 event with the same eight stressor items and the same factor structure.

Momentary mood was measured at events A1 to A6 by 14 mood adjective items with two subscales (*Distress* and *Positive Mood)* derived from an exploratory factor analysis (see section 2.2.3.2.3.4 of Chapter 2). Mood items were preceded by the words "Right now I feel…" and VAS anchor points were "Disagree" (0) and "Agree" (10). A "Not applicable" response was not available for mood items. The momentary fatigue item was not relevant to this chapter.

4.2.4 Biochemical analysis

Saliva samples were sent in one batch to the Biochemical Laboratory at the Division of Theoretical and Clinical Psychobiology, University of Trier, Germany, where they were analysed for cortisol content using a time-resolved immunoassay with fluorescent detection (Dressendörfer et al., 1992). Each sample was measured in duplicate, with an intra-assay coefficient of variance between 4.0% and 6.7%, and inter-assay coefficient of variance between 7.1% and 9.0%. The detection limit for the assay was 0.173 nmol/L.

4.2.5 Statistical analysis

Multilevel modelling (MLM) was used to examine CSA. Traditionally, studies have examined between-individual differences in CSA, assuming these differences are stable over time and eliminating within-subject variability. MLM permits the simultaneous modelling of both within- and between-individual variance, and accounts for the non-independence of observations (intra-class correlation) commonplace in nested designs. MLMs are robust against missing observations and can model heteroscedasticity if present within the data (S. Liu et al., 2012; Singer & Willett, 2003). In all models examining group comparisons, group (control, RRMS) was indicated by a dummy variable (0, 1)

at the individual level (Level-2 or Level-3, depending on the model), and comparisons were made by including a group by "effect of interest" interaction in the model. Model specifications use notation from Snijders and Bosker (2012). All analyses were conducted using SPSS Version 20.0.

4.2.5.1 Non-compliance and outliers

The *a priori* criteria for dealing with missing cortisol samples were presented in section 2.2.6.4 of Chapter 2. Before momentary associations were tested (stress response hypothesis), data cleaning was carried out as per recommendations by McCabe, Mach, and Fleeson (2011). These criteria led to the removal of unreliable responses (>90% identical responses to items within an event, n = 7), and slowly-completed events (> mean time + 3*SD*s, n = 18) which threatened the validity of momentary associations.

Potential cortisol outliers were identified using the 3*SD*s from the mean criterion (Schlotz, 2011), and observed in figures to assess the likelihood that they were influential. Figures A-E in Appendix N present scatterplots of CAR measures and transformed daytime cortisol assessments (A1-A6) against participant ID. Outliers leading to somewhat skewed model residual distributions, thus risking violation of the assumption of normality in residuals, were removed in sensitivity analyses. Residual distributions for all models are presented in Appendix O.

4.2.5.2 Cortisol awakening response

To test the hypothesis that the CAR was elevated in RRMS, a set of separate 2-level multilevel models were built nesting four daily assessments (Level-1) within individuals (Level-2) for each of the three outcomes: AUCg, AUCi, and S1 cortisol. There is a lack of consensus in the literature as to how to compute valid effect sizes from multilevel parameters that are comparable to traditional effect size estimates, such as Cohen's *d* (Peugh, 2010; Snijders & Bosker, 2012). Therefore, independent samples t-tests were used to demonstrate differences using traditional statistical analyses of means and *SD*s and Cohen's *d* effect sizes computed to facilitate comparisons with previous research.

Equation 5 specifies the CAR null multilevel model, before potential covariates or group differences were considered. Raw S1 cortisol values

showed a strong positive skew and a natural log-transformation brought the distribution of model residuals closer to normal (see Appendix O). This procedure was not necessary for CAR AUCg and AUCi, which were not substantially non-normal (see Appendix O).

Equation 5

$$CAR_{di} = \gamma_{00} + U_{0i} + R_{di}$$

Where $R_{di} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$

Where CAR_{di} is the value of the CAR outcome (AUCg, AUCi, or S1) for person i on day d. The average intercept is indicated by γ_{00}. Random effects at Level–2 (individual) are indicated by U with U_{0i} denoting the deviation of the intercept for person i from the average intercept. R_{di} indicates the residuals for each unique assessment.

Each Level-1 unit was equally distributed in time (1 day), so different error variance-covariance matrices were examined using the null model to determine model of best fit (AIC and BIC criterion), as suggested by Singer and Willett (2003). The compound symmetry matrix was chosen to model CAR AUCg and S1 cortisol, and the first order autoregressive matrix for the CAR AUCi model (see Appendix P).

Potential covariates were tested using the stepwise method suggested by Hox (2010), adding Level-1 covariates in a first step, and then Level-2 covariates, with only significant covariates retained (α = .10 criterion used for covariates) before moving to the next step. All Level-1 covariates were centred to the person-mean. Covariates were added where there were theoretical or intuitive reasons to do so: Level-1 covariates sleep quality, awakening time, and morning stress, and Level-2 covariates age, gender, smoker-status, oral contraceptive use, menstrual phase, and DMTs were examined. No covariates were retained.

It has been argued that CARs not demonstrating a cortisol increase after awakening may indicate non-adherence by participants (Thorn et al., 2006). Although Thorn et al. (2006) suggest any rise (> 0) in cortisol level post-awakening demonstrated adherence, others have suggested a 2.5nmol/L increase as demonstrating a response (Wüst et al., 2000). However, neither

criterion was based on empirical evidence. Recently, R. Miller, Plessow, Kirschbaum, and Stalder (2013) demonstrated that a 1.5nmol/L criterion is optimal when distinguishing responders from non-responders, simultaneously minimising false-positives while increasing true-positives. Analyses were therefore replicated with responders only using the 1.5 nmol/L criterion.

4.2.5.3 Diurnal cortisol slope

To test the hypothesis that daytime cortisol is elevated in RRMS, and that DCSs are flatter, a multilevel linear growth model was built which nested six (A1-A6) assessments (Level-1) within 4 days (Level-2) within 76 individuals (Level-3). DCS was operationalized as the linear slope trajectory, which is appropriate where using a variable-occasion assessment design. Raw cortisol values were transformed using the natural log-transformation to reduce the positive skew of the model residual distribution (see Figure B in Appendix O).

Equation 6 specifies the linear growth model for transformed cortisol, before covariates were considered. To facilitate interpretation of model intercepts, time was centred at 1000h (*CentredTime = Time of Assessment (hours) – 10*) meaning the intercept represented mean transformed cortisol at 1000h. Equation 6 contains the fixed effect of *CentredTime*, which was permitted to vary between individuals, where there was significant variability to model; there was no statistically significant variability in the *CentredTime* effect between-days, so this random effect was discarded in favour of a more parsimonious model. An unstructured variance-covariance matrix was used to get an estimation of the covariance of intercept and time random effects.

Equation 6

Random linear growth model for natural logarithm transformed cortisol

$$ln(CORTISOL)_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$ and $\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & \sigma_{0010} \\ \sigma_{1000} & \sigma_{10}^2 \end{bmatrix}\right)$

Where $\ln(CORTISOL)_{adi}$ is the value of natural logarithm transformed cortisol for person i on day d at assessment a; $TIME_{adi}$ the time of day (centred at 1000h) for person i on day d at assessment a. Fixed effects are denoted by γ, with γ_{000} indicating the average intercept, and γ_{100} the average effect of *TIME*. Random effects at Level–3 (individual) are indicated by V, with V_{00i} denoting the deviation of intercept for person i from the average intercept, and V_{10i} denoting the deviation of the effect of person i from the average effect of *TIME*. Random effects at Level–2 (day) are indicated by U, with U_{0di} denoting the deviation of intercept for day d from the average intercept of person i. R_{adi} indicates residuals at the level of the individual assessments.

Level-1 covariates eating a meal, drinking coffee, smoking, being at work, and physical exertion; Level-2 covariates awakening time, sleep quality; and Level-3 covariates menstrual phase, oral contraceptive use, age, gender, chronic stress, and DMT were examined in the same stepwise covariate procedure. Recent meal eating and smoking were retained, and were added as fixed effects to the model specified in Equation 6.

4.2.5.4 Cortisol reactivity

A multilevel growth model based on the same 3-level data structure was built. Transformed cortisol values and *CentredTime* was again used. Previous studies examining within-subjects effects on cortisol have used fourth polynomial growth parameters when estimating baseline cortisol growth curves (Jacobs et al., 2007; Peeters et al., 2003). Therefore, models were specified with consecutive polynomial time parameters up to a fourth polynomial (see specifications in Appendix Q). As with the linear model, all models specified random growth effects at Level-3 (between individuals) only, and this random effect represented the linear effect only for parsimony. Table 15 reports

parameters for each baseline model (random linear, random quadratic, random third polynomial, random fourth polynomial) along with deviance statistics to ascertain the most efficacious, parsimonious model. The deviance statistic is a chi-squared value equalling the difference in -2 log likelihood estimates between models, with degrees of freedom equalling the difference in the number of model parameters estimated (Snijders & Bosker, 2012). A statistically significant deviance statistic indicates a significant improvement in data modelling. Table 15 shows that the random fourth polynomial was the most efficacious model, not yet considering covariates (Deviance Statistic versus random linear, $\chi^2 = 29.56$, $df = 3$, $p < .0001$; versus random quadratic, $\chi^2 = 12.20$, $df = 2$, $p < .001$). The baseline model specification is presented in Equation 7 below, and rejected growth model specifications are detailed in Appendix Q.

Equation 7

Model 4: Random fourth polynomial growth model

$$\ln(CORTISOL)_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + \gamma_{300}(TIME_{adi})^3 + \gamma_{400}(TIME_{adi})^4 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$ and $\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & \sigma_{0010} \\ \sigma_{1000} & \sigma_{10}^2 \end{bmatrix}\right)$

Where $\ln(CORTISOL)_{adi}$ is the value of natural log transformed cortisol for person i on day d at assessment a; $TIME_{adi}$ the time of day (centred at 1000h) for person i on day d at assessment a. Fixed effects are denoted by γ, with γ_{000} indicating the average intercept, γ_{100} the average effect of *TIME*, γ_{200} the average effect of $TIME^2$ (quadratic), γ_{300} the average effect of $TIME^3$, γ_{400} the average effect of $TIME^4$. Random effects at Level-3 (individual) are indicated by V, with V_{00i} denoting the deviation of intercept for person i from the average intercept, and V_{10i} denoting the deviation of the effect of person i from the average effect of *TIME*. Random effects at Level-2 (day) are indicated by U, with U_{0di} denoting the deviation of intercept for day d from the average intercept of person i. R_{adi} indicates residuals at the level of the individual assessments.

Table 15. *Building the Baseline Multilevel Model for testing the Within-Subject Contextual Effects of Stress and Mood on Natural-log Transformed Cortisol*

	Parameter		Model 1 Random Linear[a]	Model 2 Fixed Quadratic Random Linear[a]	Model 3 Fixed 3rd Polynomial Random Linear[a]	Model 4 Fixed 4th Polynomial Random Linear[a]	Model 5 Fixed 4th Polynomial Random Linear[a] Fixed covariate
	Fixed Effects						
Composite Model	Intercept (ln cort at 1000h)	γ_{000}	1.7621***	1.6484***	1.6626***	1.9092***	1.8945***
	TIME	γ_{100}	-0.0927***	-0.0300~	-0.0437	-0.4024***	-0.3922***
	$TIME^2$	γ_{200}		-0.0064***	-0.0031	0.1484***	0.1380**
	$TIME^3$	γ_{300}			-0.0002	-0.0241***	-0.0218**
	$TIME^4$	γ_{400}				0.0013***	0.0011**
	MEAL	γ_{500}					0.0785***
	SMOKE	γ_{600}					0.1898*
	Random parameters/variance components						
Level-1	Within-day	σ^2_R	0.1149***	0.1132***	0.1132***	0.1122***	0.1102***
Level-2	Between-day	σ^2_0	0.0065*	0.0070*	0.0070*	0.0070*	0.0080**
Level-3	In intercept	σ^2_{00}	0.0821***	0.0817***	0.0816***	0.0828***	0.0812***
	Linear term (TIME)						
	Variance	σ^2_{10}	0.0010**	0.0010**	0.0010**	0.0010**	0.0011***
	Covar with intercept	σ_{0010}	-0.0042*	-0.0042*	-0.0042*	-0.0044**	-0.0045*
Goodness of fit	-2 Log Likelihood		1409.21	1391.85	1391.74	1379.65	1361.43

Note. MEAL = reported eating a meal in prior 30 minutes; SMOKE = reported smoking in the last 30 minutes.

[a] Random linear only at the individual level (Level-3).

*** $p < .001$, ** $p < .01$, * $p < .05$.

4.2.5.4.1 Mediation analysis

In an exploratory analysis, MLM was used to test whether distress mediated the relationship between stressors and cortisol using a mediation procedure described by Jacobs et al. (2007). First, it was determined whether Stressor Hindrance and/or Stressor Challenge were associated with a Distress response by entering both with a time predictor as fixed effects in a multilevel model with Distress score as outcome. A statistically significant association would establish a relationship between the predictor and mediator. Then both Distress and Stressor variables were entered into a multilevel model with transformed cortisol as outcome, with partial mediation indicated by a smaller stress-cortisol relationship and total mediation if the effect was eliminated. The mediation analysis was performed using all participant data, and then the RRMS group only.

4.3 Results

4.3.1 Cortisol awakening response

The CAR AUCg and AUCi analyses were based on 258 assessment days (Level-1) nested within 75 participants (37 RRMS; 38 control). S1 cortisol was based on 298 assessment days (Level-1) nested within 76 participants (38 RRMS; 38 control). Overall, 881 S1, S2, and S3 samples were provided out of a possible 912 (96.6%), with similar compliance rates in the RRMS group (96.5%) and control group (96.7%). Timestamps indicated six of the S1 samples were provided 10 minutes after 0830h, 13 S2 samples more than 37 minutes after awakening (S1), and 22 S3 samples more than 52 minutes after awakening (S1). Valid CAR computations, requiring all three valid samples, were therefore possible on 258 out of 304 (84.8%) assessment days. In the RRMS group, valid CARs were provided on all 4 days in 63.2% of participants, at least 3 days in 89.5%, and at least 2 in 97.4%. In the control group, all 4 days were valid in 57.9% of participants, at least 3 in 73.6%, and at least 2 in 100%. One RRMS participant was excluded from the CAR dataset as they provided no compliant CARs. No participants were excluded based on missingness for the S1 analysis. There were 177 out of 258 (68.6%) CAR measurement days classified as "responder" days, with similar prevalence in the RRMS group (69.7%) and the control group (67.5%).

Figure 9A presents the CAR for each group represented by the mean of within-subjects means for salivary cortisol levels at S1, S2, and S3. Mean cortisol levels were higher in the RRMS group at every assessment and suggested minimal difference in the dynamic of the CAR, with relatively parallel trends; particularly in "responders" only (Figure 9B).

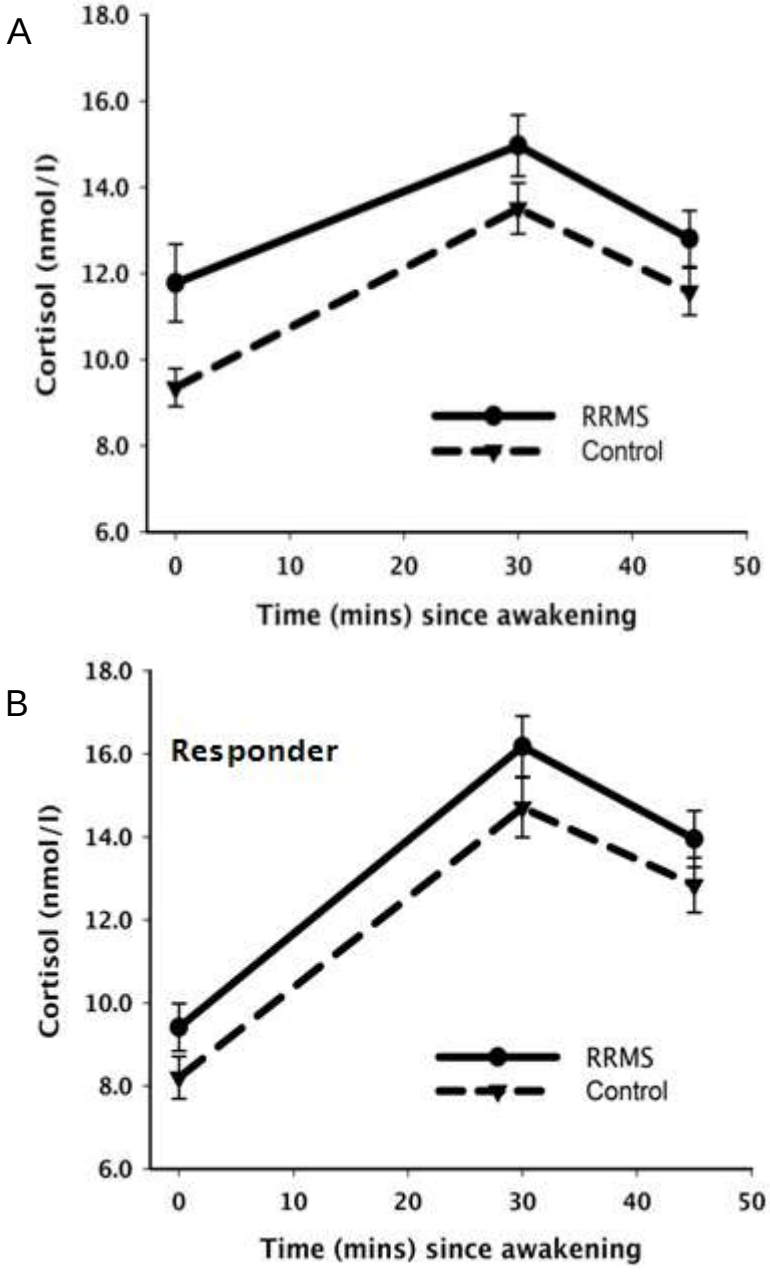


Figure 9. Cortisol awakening response represented by the mean of the within-subject means for samples at 0 (S1), 30 (S2), and 45 minutes (S3) post-awakening. (A) includes responder and non-responder data; (B) represents responders-only on 1.5nmol/L criterion. Error bars represented the standard error of the mean.

A summary of findings for the group comparisons for the CAR is provided in Table 16, which presents untransformed cortisol data for ease of interpretation. A statistically significant difference between groups was found for CAR AUCg, p = .028, 95% CI [0.37, 6.49], such that, on average, AUCg was 1.71 nmol/L/min greater in the RRMS group than control group (Figure 10). The random effects portion of the model revealed 3.4% of CAR AUCg variability could be attributed to group (calculation from Snijders & Bosker, 2012, p.112). Removing the possible outlying assessment from the dataset (see Appendix N) led to similar results, p = .044, 95% CI [0.09, 6.00]. However, with "responders" only, the main effect was diminished, p = .147, 95% CI [-0.41, 2.66]. When re-running the model (all data) with a depressive symptoms fixed effect as a potential moderator, the group effect retained a trend towards statistical significance, p = .081, 95% CI [-0.18, 3.14], and depressive symptoms was not a statistically significant moderator, p = .49, 95% CI [-0.22, 0.46].

Table 16 Cortisol Awakening Response Group Difference Multilevel Model Parameters

	RRMS γ (*SE*)	Control γ (*SE*)	Group difference γ (*SE*)
CAR AUCg	13.55 (0.54)	11.83 (0.54)	1.71 (0.77)*
CAR AUCi	1.78 (0.47)	2.38 (0.48)	-0.59 (0.67)
S1 cortisol	11.86 (0.69)	9.38 (0.69)	2.49 (0.98)~
S2 cortisol	14.93 (0.64)	13.47 (0.64)	1.46 (0.91)
S3 cortisol	12.72 (0.58)	11.59 (0.58)	1.13 (0.82)

Note. CAR AUCg indicates cortisol awakening response area under the curve ground; CAR AUCi, cortisol awakening response area under the curve increase; S1, cortisol upon awakening; S2, cortisol 30min after awakening; S3, cortisol 45min after awakening. Untransformed S1, S2, and S3 cortisol parameters presented in table, but statistical significance based on transformed data. Data presented is from responder and non-responder days.

*p < .05, ~ p = .07.

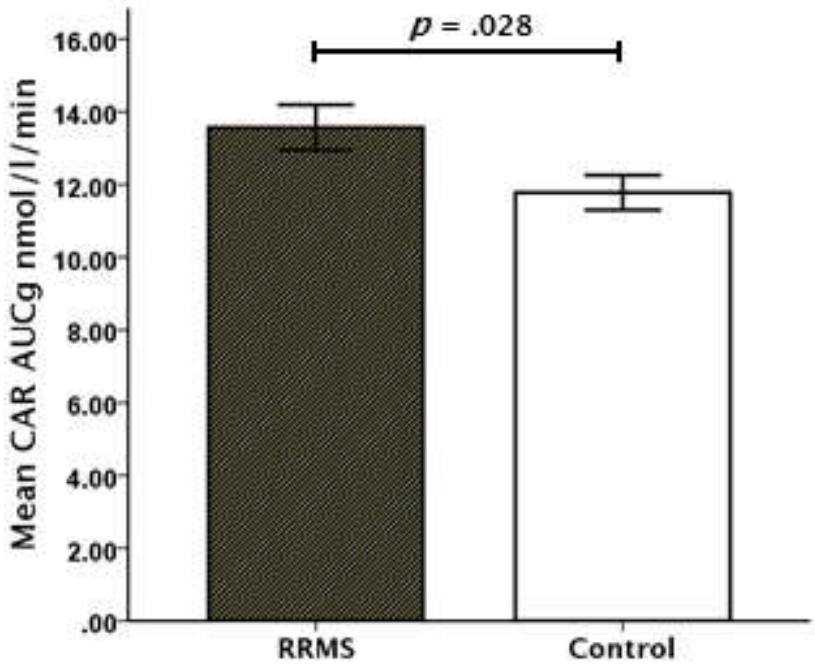


Figure 10. Bar chart of cortisol awakening response AUCg for the RRMS and control groups. Error bars indicate standard error of the mean.

There was no effect of group in the CAR AUCi model, p = .38, 95% CI [-1.92, 0.74] (see Figure 11). Four possible outliers were identified (see Appendix N, Figure B), but their removal did not significantly change the result, p = .82, 95% CI [-1.36, 1.07]. "Responder"-only analysis also showed no group effect, p = .99, 95% CI [-0.98, 0.98].

Although there appeared a difference in S1 cortisol in Figure 9A, this represented only a trend towards a statistically significant difference between groups, p = .069, 95% CI [-0.02, 0.31]. Figure 12 presents S1 cortisol differences between groups using raw untransformed data. "Responder"-only analyses revealed no difference in S1 cortisol between groups, p = .16, 95% CI [-0.04, 0.26]. Depression led to a reduction in the group effect for S1 transformed cortisol, γ = .09, p = .31, 95% CI [-0.09, 0.27, but was not a statistically significant moderator itself, p = .32, 95% CI [-0.02, 0.05].

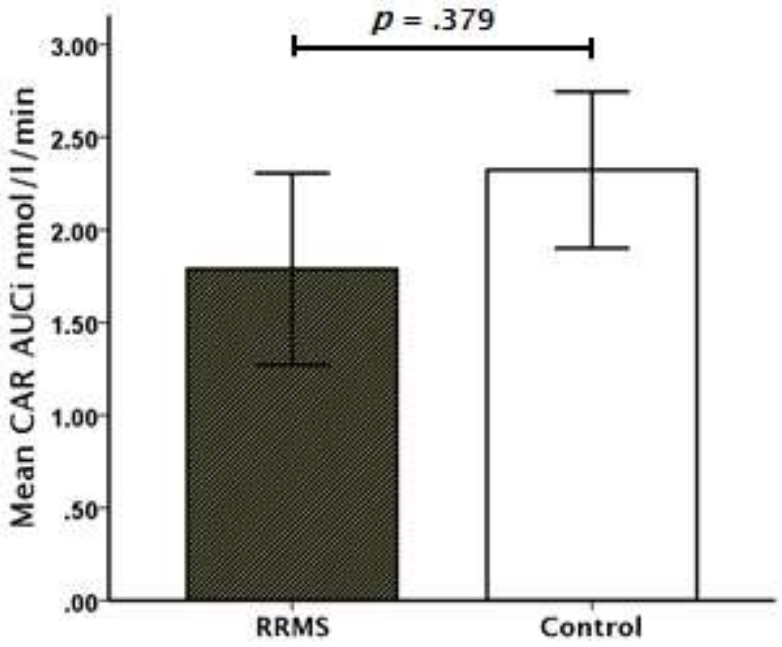


Figure 11. Bar chart of cortisol awakening response AUCi for the RRMS and control groups. Error bars indicate standard error of the mean.

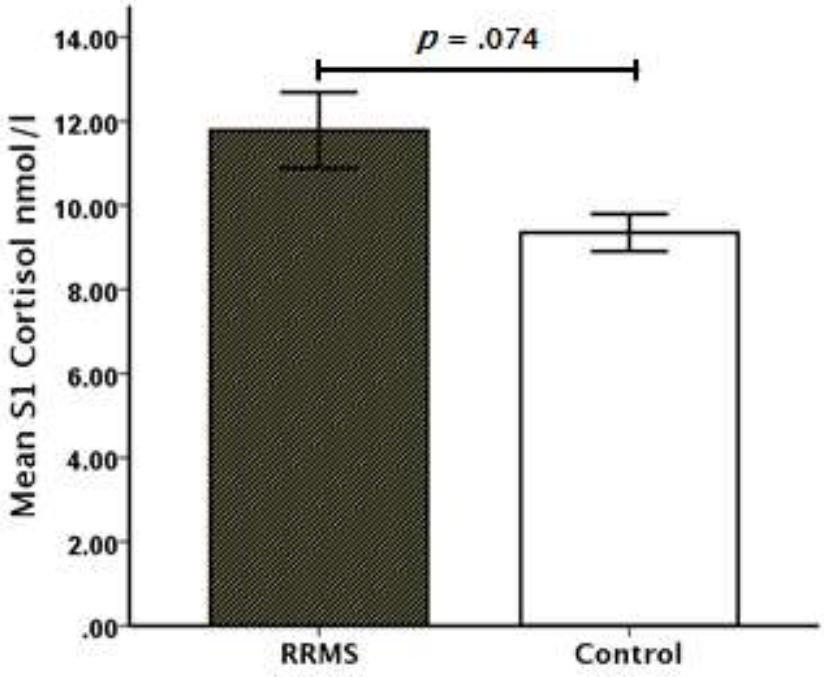


Figure 12. Bar chart of untransformed S1 cortisol for RRMS and control groups. Error bars indicate standard error of the mean.

Note. p-value represents result using transformed data.

Table 17 presents group comparisons conducted on within-subjects means. Results demonstrated a statistically significant main effect of group for CAR AUCg (t = 2.29, p = .025), with a medium effect size (d = 0.53). There was a similarly elevated cortisol level in the RRMS group for each assessment (S1-S3), albeit not to statistical significance. There was no group difference for CAR AUCi.

Table 17 Mean of Within-Subjects Means (*SD*) for CAR Outcomes in each group

CAR measure	RRMS	Control	*t*	*p*	95% CI	Cohen's *d*
AUCg	13.57 (3.77)	11.78 (2.95)	2.29	.025	[0.23, 3.35]	0.53
AUCi	1.88 (3.22)	2.02 (2.69)	0.21	.840	[-1.51, 1.22]	-0.04
LN S1 cort	2.41 (0.41)	2.28 (0.31)	1.49	.140	[-0.04, 0.29]	0.36
LN S2 cort	2.71 (0.28)	2.60 (0.26)	1.74	.086	[-0.02, 0.23]	0.41
LN S3 cort	2.56 (0.29)	2.46 (0.28)	1.42	.210	[-0.04, 0.23]	0.35

Note. "Responder" and "non-responder" data included. Significance tests use independent samples t-test.

4.3.2 Diurnal cortisol slope

Compliance with the quasi-random design was excellent, with 1661 out of 1824 (91.1%) of events completed. Some missing salivettes or missing/incorrect random codes left 1637 out of 1661 (98.6%) samples (97.9% in RRMS; 99.2% in controls) to analyse. Eight assessment days belonging to eight individuals (five RRMS; three control) contained less than three cortisol assessments, excluding those assessment days from the analysis. All participants provided more than 2 valid assessment days.

Figure 13 presents fitted linear regression lines for each participant in the RRMS and control groups, along with thick fitted lines representing the mean average slope for each group. The linear time model, not yet including the group predictor, accounted for 22.0% of the total variance in transformed cortisol. Intra-class correlation coefficients revealed 55.0% of overall cortisol variability remaining was partitioned within days, 4.5% between days, and 40.5% between individuals: approximately 60% of variability in cortisol was within individuals, and 40% between individuals. Figure 13 suggested minimal difference in cortisol slope between groups (see Appendix R for average loess curves for each group), and individual DCS were relatively homogeneous within groups, although there remained some variability. In both groups, there was a statistically significant decreasing trend in cortisol levels over time (see Table 18), although this average linear trend did not differ between groups, p = .61, 95% CI [-0.025, 0.015]. There was no group difference in mean log-transformed cortisol levels between groups at 1000h, p = .15, 95% CI [-0.039, 0.257]. Centring the intercept at 2000h instead of 1000h revealed no difference between groups in evening levels of cortisol, γ = 0.058, SE = 0.054, p = .48, 95% CI [-0.106, 0.222]. Adding group and group by time fixed effects to the model reduced variability in transformed cortisol by 1.51%.

Table 18 Multilevel Model Parameter Estimates for Natural-log Transformed Cortisol as a Function of Time and Group

	RRMS γ (SE)	Control γ (SE)	Group difference γ (SE)
Intercept (1000h)	1.797 (0.053)***	1.688 (0.053)***	0.109 (0.074)
Time (per hour)	-0.096 (0.007)***	-0.091 (0.007)***	-0.005 (0.010)

Note. Fixed effects of covariates (recent meal, recent smoking), and random effects parameters are not presented.

*** p < .001

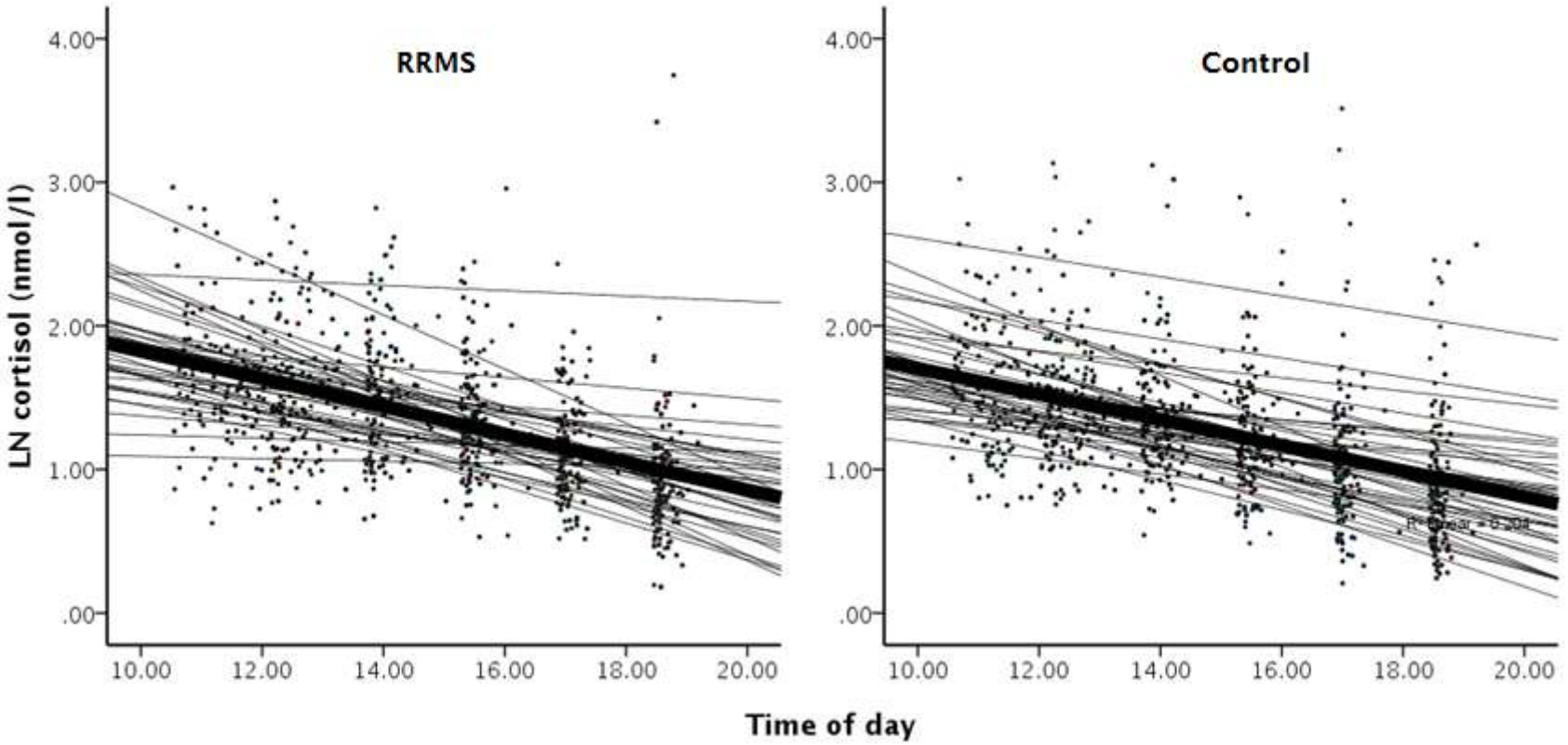


Figure 13. Spaghetti plots of average group (thick) and individual-specific (thin) regression lines for natural log-transformed cortisol as a function of time of day. Small black markers indicate individual salivary assessments.

4.3.3 Within-subjects contextual stress and mood reactivity.

There were seven momentary events (0.4%) where participants provided identical responses to more than 90% of presented items, and 18 events (1.0%) with slow completion of events such that it took longer than 6 minutes 46 seconds (mean + 3*SD*s) to complete. A histogram of event completion speeds to demonstrate problematic assessments is presented in Appendix I.

Table 19 summarises tests of associations of stressors and momentary mood with cortisol levels. Across all groups, Stressor Hindrance since the last event was positively associated with subsequent momentary cortisol, $p = .030$, 95% CI [0.001, 0.030], meaning that for every 1-unit increase above an individual's usual level of Stressor Hindrance, there was an average 0.016 unit increase in log-transformed cortisol (1.61% increase in cortisol) from its diurnal pattern. This effect was not statistically significant in either the RRMS group, $p = .093$, 95% CI [-0.003, 0.035], or the control group, $p = 0.166$, 95% CI [-0.006, 0.037], probably due to loss of power with smaller number of observations. There was no difference in the Stressor Hindrance effect between groups, $p = .944$, 95% CI [-0.030, 0.028].

Across all groups, prior Stressor Challenge was associated with a decrease in momentary cortisol, $\gamma = -0.010$, $p = .033$, 95% CI [-0.020, -0.001], describing an average 0.010 unit decrease in log-transformed cortisol (-1.00 % decrease in cortisol) from the diurnal pattern for every 1-unit increase in Stressor Challenge above the usual level. This effect was not present in either group, although the control group demonstrated a trend towards statistical significance: RRMS group, $p = .192$, 95% CI [-0.021, 0.004]; Control group, $p = .082$, 95% CI [-0.026, 0.002].

Momentary Distress was associated with an increase in momentary log-transformed cortisol level, $p = .019$, 95% CI [0.003, 0.033], such that a 1-unit increase in Distress was associated with an increase of 0.018 units in log-transformed cortisol (1.82% increase in cortisol). Positive Mood was not associated with cortisol, $p = .483$, 95% CI [-0.009, 0.020]. As stated in Table 19, there were no group differences for either Distress or Positive Mood.

Table 19 Multilevel Model Parameters for Effect of Daily Stressors and Associations of Momentary Mood with Natural-Log Transformed Cortisol

	All groups	RRMS	Control	Group difference
	γ (*SE*)	γ (*SE*)	γ (*SE*)	γ (*SE*)
Stressor				
Hindrance	0.016 (0.007)*	0.016 (0.010)~	0.015 (0.010)	0.001 (0.015)
Challenge	-0.010 (0.005)*	-0.008 (0.006)	-0.012 (0.006)~	-0.003 (0.009)
Mood				
Distress	0.018 (0.008)*	0.019 (0.010)~	0.016 (0.012)	0.003 (0.015)
Positive Mood	0.005 (0.007)	0.003 (0.010)	0.008 (0.011)	-0.005 (0.015)

Note. Intercept, time parameters (up to fourth polynomial), fixed effects of covariates (recent meal, recent smoking), and random effects parameters are not presented.

* $p < .05$, ~ $p < .010$.

4.3.3.1 Mediation of stressor-cortisol association

A 1-unit increase in Stressor Hindrance above an individual's usual level was associated with a 0.38 unit increase in Distress, γ = 0.38, $p < .001$, 95% CI [0.34, 0.42]. However, there was no association between Stressor Challenge and Momentary Distress, γ = 0.01, $p = .56$, 95% CI [-0.02, 0.04]. There was no statistically significant Group by Stressor Hindrance interaction, γ = -0.01, $p = .79$, 95% CI [-0.09, 0.07], meaning this effect was not different between the groups. Since Stressor Challenge was not associated with Distress, this measure was excluded from further mediation tests.

The mediation pathways are presented in Figure 14. As can be seen, the mediation pathway has been established from the findings already presented, and that once Distress was included in the model with Stressor Hindrance in predicting transformed cortisol, the effect of Stressor Hindrance had diminished, γ = 0.008, $p = .355$, 95% CI [-0.009, 0.024]. In the RRMS group only (lower panel below) the parameters are very similar to the mediation model with all participants, but some relationships fail to reach statistical significance due to a smaller *n* and reduced statistical power.

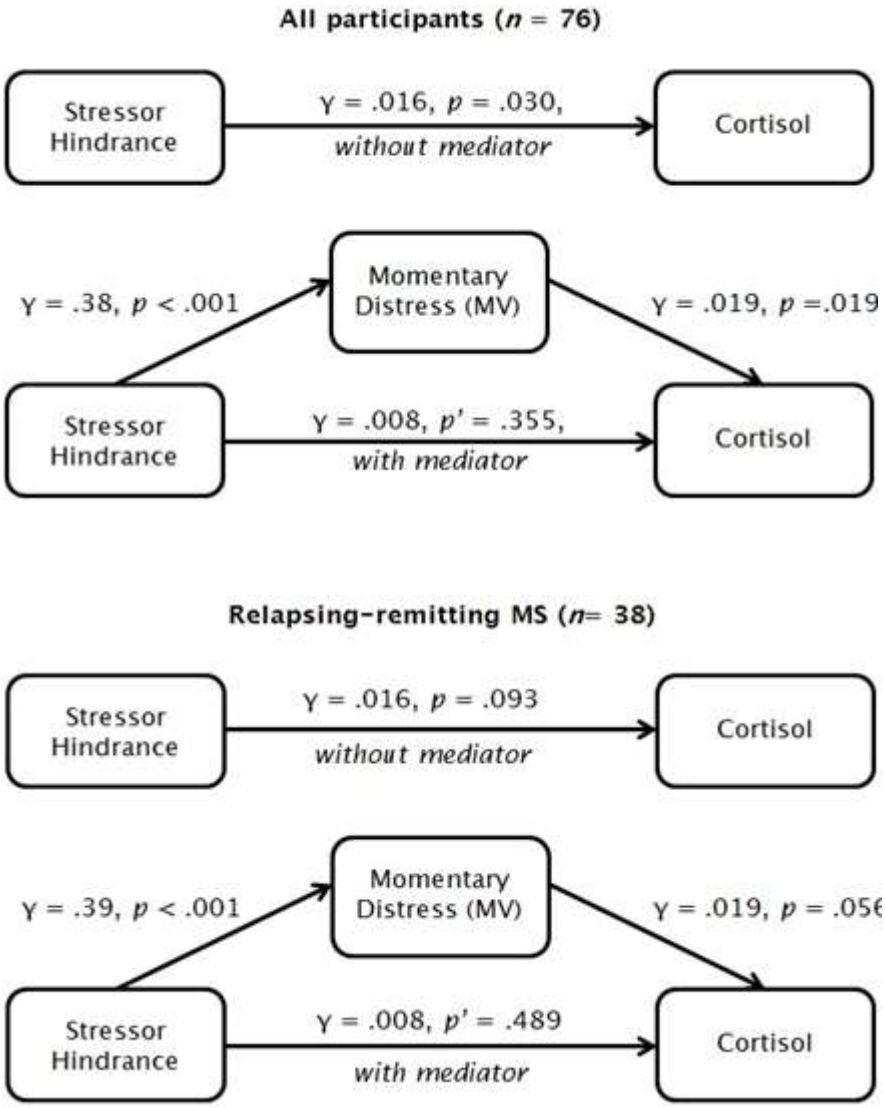


Figure 14. Mediation pathways demonstrating the mediating effect of momentary distress in the stressor hindrance – cortisol relationship.

Note. Transformed cortisol used in all analyses.

4.3.4 Results summary

CAR output (AUCg) was greater in the RRMS group compared to the control group. S1 cortisol was also typically elevated in people with RRMS compared to controls, but not to statistical significance. Depressive symptoms did not moderate group differences in CAR measures. The DCS was no different between groups. Cortisol reactivity to daily life stressors was present in both groups, mediated by the distress response.

4.4 Discussion

The study was designed to explore CSA in everyday life in a well-defined RRMS group (in remission-phase) with no comorbidities. It was demonstrated that increased cortisol secretory activity (HPA axis hyperactivity) is evident in the first 45 minutes after awakening in an RRMS group compared to a healthy control group. However, there was no evidence of flatter DCS or elevated cortisol levels in the evening (2000h) in RRMS.

Cortisol reactivity to everyday stressors was present in both groups, albeit in a positive direction for Stressor Hindrance and, curiously, a negative direction for Stressor Challenge. However, Distress responses were present only in response to Stressor Hindrance, and not Stressor Challenge, with the Distress response mediating cortisol reactivity to Stressor Hindrance in both groups. The results did not show any group differences for any of the within-subjects tests, indicating individuals with RRMS demonstrated no dysfunction in emotional or HPA axis reactivity to everyday life stressors.

4.4.1 HPA axis hyperactivity

The hypothesis that CARs would be larger in RRMS than controls was supported by the study, supporting previous findings of HPA axis hyperactivity in RRMS (Gold et al., 2010; Kern et al., 2013; Kern et al., 2011; Michelson et al., 1994; Ysrraelit et al., 2008). However, while several previous studies argued depressive comorbidity was the main source of HPA axis hyperactivity in RRMS (Gold et al., 2010; Gold et al., 2011; Kern et al., 2011), the present study found no support for the hypothesised moderating role of depressive symptoms in HPA axis hyperactivity in MS. Findings were in concordance with those reported by Kern et al. (2013) where salivary cortisol levels were not

associated with depressive symptoms in MS, and supports other studies finding no association between depressive symptoms and serum cortisol (Ysrraelit et al., 2008) and cortisol in post-mortem cerebrospinal fluid (Melief et al., 2013). It should be noted that results here do not rule out that MDD comorbidity would further elevate the CAR output beyond that which is associated with RRMS.

The finding of increased total cortisol output in the CAR in RRMS concurs with both existing case-control studies specifically examining the CAR (Kern et al., 2013; Kern et al., 2011). The group comparison effect size (d = 0.53) was slightly lower than that found by Kern et al. (2011) (d = 0.80) but very similar to Kern et al. (2013) (d = 0.58) who also recruited no individuals with "clinically significant" depression to their sample. In addition, the difference between the RRMS BDI-low group and healthy control group in the Kern et al. (2011) study also demonstrated a medium effect size (d = 0.55), albeit not statistically significant (likely due to low power).

Different CAR computations from constituent assessments provide specific information about the nature of the CAR (Chida & Steptoe, 2009; Clow et al., 2004; Fekedulegn et al., 2007). The present study provides the first empirical examination of the S1 cortisol level, which is thought to measure the end of the pre-awakening cortisol rise (Clow et al., 2010). Although no statistically significant difference was found between groups for S1 cortisol, there was a clear trend towards higher levels in RRMS. The results suggest hyper-secretion of cortisol in RRMS likely occurs pre-awakening with "normal" (similar to controls) awakening responses (CAR AUCi) occurring on top of this elevated waking cortisol level: a combination resulting in greater CAR AUCg levels in the RRMS group.

It cannot explicitly be discerned from either Kern et al. (2011) or Kern et al. (2013) whether CAR AUCi was elevated in respective RRMS groups as this measure was not computed, and nor was S1 cortisol examined in statistical analyses. However, it is possible to examine the CAR plots in each study to scrutinize what was observed. The Kern et al. (2013) study's CAR plot (p.6) demonstrated a similar pattern of results to that presented here: high cortisol levels seemingly present upon waking in RRMS, and similar group differences at T0, T30, T45, and T60 suggesting no group differences in CAR AUCi. Conversely, the CAR plot from the earlier study (Kern et al., 2011, p.1508)

displayed no clear difference in waking cortisol levels in RRMS and controls, but a pronounced CAR AUCi in RRMS which may or may not have reached statistical significance. Although speculative, it demonstrates heterogeneity in findings despite seemingly similar results. One cause for this heterogeneity may be methodological factors and, specifically, a lack of control over the sampling protocol in everyday life. This is a particular strength of the present study, discussed in more detail in section 4.4.3 below.

When examining "responders" only, group differences in CAR AUCg and S1 were eliminated, which was potentially an important observation as non-response is often held as indicative of non-compliance in CAR research. In two studies where participants were monitored with polysomnography (objective observation of awakening), 100% of participants demonstrated a response to awakening (Gribbin, Watamura, Cairns, Harsh, & LeBourgeois, 2012; Wilhelm et al., 2007). Wilhelm et al. (2007) argued that awakening in a laboratory may represent a more arousing event than usual waking at home, which may contribute to higher responder-rates in the laboratory than everyday life. It remains possible that "non-response" is, in itself, a feature of some CARs (Clow et al., 2004).

Clow et al. (2010) have described a period of dissociation between ACTH and cortisol in the pre-awakening period, such that adrenal sensitivity to ACTH is attenuated prior to awakening in healthy individuals. Evidence supporting this claim was provided by studies examining both ACTH and cortisol in blood both before and after awakening: ACTH levels increased far more quickly than cortisol in the pre-awakening period (D. H. Hellhammer et al., 2009; Wilhelm et al., 2007). The mechanism underlying this dissociation is unclear. It has been suggested inhibitory actions of the hippocampus on cortisol secretion are most prominent prior to awakening when rapid eye movement sleep is dominant and hippocampal activity is heightened (Clow et al., 2010). Although results are mixed in the literature, hippocampal damage has previously been shown to have an inverse relationship with waking cortisol levels such that smaller hippocampal volumes were associated with higher waking levels (Beresford et al., 2006; Frodl & O'Keane, 2013; O'Hara et al., 2007). It is also noticeable in a study by Buchanen et al. (2004) that S1 cortisol level was elevated in individuals with hippocampal damage. Average hippocampal volumes have been shown to be reduced in pwRRMS (12% smaller) compared to

healthy individuals (Sicotte et al., 2008) so one could speculate that reduced hippocampal functioning in RRMS may produce elevated S1 cortisol via a reduction in typical ACTH-cortisol dissociations prior to awakening. Gold et al. (2010) have shown reduced hippocampal volumes in an RRMS group compared to controls, but did not find a corresponding elevation in S1 cortisol to support this hypothesis. The hypothesis is beyond the scope of this thesis, but is worth further exploration with strategies to maximise sampling protocol compliance.

In contrast to earlier findings, there was no evidence of flatter DCS or elevated evening cortisol levels in RRMS. In previous studies, flatter DCS and elevated evening cortisol were demonstrated in an RRMS group with high levels of depressive symptoms (≥14 on the BDI-II) (Gold et al., 2010) and in an MS group (RRMS and SPMS) where depressive symptoms were not measured (Michelson et al., 1994). Gold et al. (2011) also reported flatter DCS and elevated evening cortisol level in an RRMS group with comorbid MDD compared to RRMS without MDD. Other studies have failed to observe elevated evening cortisol levels at 2200h (Kern et al., 2013; Kern et al., 2011), but have not included many participants with high levels of depressive symptoms. Combining these findings, it may be that depressive comorbidity is most relevant to evening cortisol levels and resulting flatter DCS rather than driving any changes in the CAR.

4.4.2 Cortisol and emotional reactivity

There was no evidence to support the hypothesis regarding attenuated cortisol reactivity to everyday stressors in RRMS. This study represents the first to demonstrate cortisol and emotional reactivity to stress in RRMS. The fact that similar cortisol reactivity to stress was reported across groups, as well as similar mechanisms (mood mediation), suggests a fully-functioning HPA axis-related stress response system in RRMS. Stressor Hindrances were associated with cortisol reactivity, which was in line with experimental studies where threat of failure typically elicits strong positive cortisol responses (Dickerson & Kemeny, 2004), and with everyday life studies where momentary performance pressure has been associated with cortisol (Schlotz et al., 2006). The effect in the present study was mediated by mood; a mechanism similar to that found by Jacobs et al. (2007) in an EMA study with healthy women, where negative

affect mediated the relationship between “activity stress” and cortisol and “physical stress” and cortisol.

One unanticipated finding was that an increase in prior Stressor Challenge was associated with a reduction in cortisol. However, given the previous assertion that mood mediates associations between stressors and cortisol, it is perhaps not surprising there was no positive cortisol response to Stressor Challenge when this stressor subscale was not associated with a distress response. Challenges have been described as eliciting problem-focussed coping strategies that can ultimately lead to the achievement of goals (Van den Broeck et al., 2010), which may explain why negative affect did not ensue.

As attenuated responses to experimental stressors have been reported previously in RRMS samples where comorbid MDD was not excluded or measured (Ackerman et al., 1996; Heesen, Schulz, et al., 2002), it is possible that attenuated cortisol reactivity to everyday life stressors may be present in pwRRMS and comorbid depression. Peeters et al. (2003) reported attenuated cortisol reactivity to everyday negative events in MDD, whereas there was strong positive cortisol reactivity in healthy participants. Future studies should explore this possibility. Peeters et al. (2003) also found no relationship between positive events and cortisol, which supports the idea that everyday stressors with positive connotations (which Stressor Challenges potentially have) elicit no HPA axis response.

4.4.3 Strengths and limitations

The study has a number of strengths to highlight. Firstly, cortisol was sampled over a 4-weekday period which provides greater measurement reliability compared to previous studies of salivary cortisol in everyday life in RRMS, all have been conducted over 2 days (Gold et al., 2010; Gold et al., 2011; Kern et al., 2013; Kern et al., 2011). Measuring on consecutive days provided methodological rigour as bias was not introduced by self-selecting sampling days. No other study in MS has controlled for weekday-weekend effects in cortisol levels (Kunz-Ebrecht et al., 2004; Schlotz et al., 2004; Thorn et al., 2006) or controlled for potentially-confounding behaviours (Chida & Steptoe, 2009).

It was important to ensure all participants were clinically stable upon taking part. All participants had sufficient time (3 months or more) to recover from any relapse and associated corticosteroid treatments before taking part. Ysrraelit et al. (2008) clearly demonstrated increased 24h urinary cortisol levels and elevated pro-inflammatory cytokine (IL-1β, IL-6, TNF-α) levels in relapse-phase RRMS compared to remission-phase RRMS, corroborating findings of heightened HPA axis activity with higher pro-inflammatory cytokine levels elsewhere (Steensberg, Fischer, Keller, Møller, & Pedersen, 2003; Sternberg, 2006). All studies in RRMS using salivary cortisol methods published in recent years have included a variant of this control: 4 weeks in all studies except Gold et al. (2010) which had a 3-month "wash out" period.

The relative homogeneity of the RRMS group recruited is both a strength and limitation. By ensuring all participants were ambulatory, participants could be more confident in their ability to complete the protocol in their everyday lives, and eliminating depression ensured no confounding effects on cortisol (it would also overlap with fatigue symptoms assessed in other chapters). However, this does limit the generalizability to the overall RRMS population.

An important limitation was the lack of an objectively observed time of awakening. Although participants were fully briefed to engage with the handheld upon awakening and to pre-set the S1 event alarm, it was not possible to be completely certain participants did either. Although this is a common limitation of everyday life studies and this study at least provided a morning "wake up" alarm facility, validating wake up times with polysomnographical readings or movement sensors (Actiwatch) would be the gold standard (Fries et al., 2009).

Finally, although momentary assessments were relatively intensive, they may not have been sufficiently frequent to capture all effects of specific stressors. For instance, cortisol peaks follow emotional states with an approximate time lag of 10-20 minutes (Schlotz et al., 2008), meaning cortisol levels may have recovered to (or be recovering towards) normal levels by the time a saliva sample was requested. Equally, cortisol levels may not yet have peaked if the stressful event had occurred very briefly prior to the momentary assessment. Either way, this limitation could only lead to an underestimation of the effect, and may also contribute to the somewhat unexpected statistically significant negative cortisol response to Stressor Challenge.

4.4.4 Conclusion

This chapter presented evidence for greater cortisol output during the CAR period in RRMS, but this did not include a more pronounced increase from awakening levels, nor was there any difference in DCS. Cortisol reactivity to everyday life stressors was similar in both pwRRMS and healthy individuals, indicating a fully-functioning HPA axis-related stress-response system in RRMS.

Chapter 5: Salivary cortisol in everyday life and its relationship with fatigue and chronic fatigue syndrome: A systematic review[1]

Cortisol secretory activity (CSA) is relevant to fatigue due to cortisol's regulatory role in energy metabolism, the high prevalence of fatigue in conditions characterised by low cortisol levels (e.g., Addison's disease), and the efficacy of low-dose corticosteroid treatments in alleviating short-term fatigue (Cleare et al., 1999; Khani & Tayek, 2001; McKenzie et al., 1998; Sapolsky et al., 2000). Chapter 5 begins by shortly reviewing the clinical importance of fatigue, both in chronic fatigue syndrome (CFS) and as a prevalent somatic experience in other clinical and nonclinical populations. A systematic review is then presented focussing on studies investigating associations between salivary cortisol in everyday life and fatigue, both in CFS and fatigue in other populations. All studies are narratively reviewed, and a subset meta-analysis is presented for case-control CFS studies.

5.1 Introduction

Fatigue is a relatively common somatic experience. Fatigue is experienced by 11.3-37.9% of the general population (Ricci, Chee, Lorandeau, & Berger, 2007; van't Leven, Zielhuis, van der Meer, Verbeek, & Bleijenberg, 2010) and considered the primary motive for 6.5% of general practitioner visits and a discreet symptom in 19% (Cullen, Kearney, & Bury, 2002). Fatigue is prevalent in many clinical populations (Swain, 2000) and is frequently associated with detrimental impacts on daily living (Hewlett et al., 2005; Lerdal, Celius, Krupp, & Dahl, 2007). CFS is thought to conceivably represent the extreme end of a fatigue continuum (Pawlikowska et al., 1994). However, some have suggested CFS may be qualitatively and quantitatively different to

[1] A version of this chapter has been published as: Powell, D. J. H., Liossi, C., Moss-Morris, R., & Schlotz, W. (2013). Unstimulated cortisol secretory activity in everyday life and its relationship with fatigue and chronic fatigue syndrome: a systematic review and subset meta-analysis. *Psychoneuroendocrinology*, *38 (11)*, 2405-2422.

chronic fatigue per se, including fatigue secondary to distinct conditions such as MS or cancer (Wessely, 2001).

5.1.1 Chronic fatigue syndrome

CFS has unknown aetiology. The most widely used definition of CFS was provided by the US Centre for Disease Control and Prevention (CDC) (Fukuda et al., 1994). In the UK, the Oxford case definition (Sharpe et al., 1991) may also be used. The CDC definition requires at least 6 consecutive months of profound and persisting fatigue that is not attributable to other medical conditions, coexisting with several other symptoms such as muscular pain and short-term memory loss. The CDC and Oxford definitions are stated in full in Appendix S.

5.1.1.1 Cortisol secretory activity in chronic fatigue syndrome

The relevance of hypocortisolism (reduced cortisol output) to disorders such as CFS has previously been hypothesised (Fries et al., 2005; Heim et al., 2000). In an influential study of HPA axis activity in CFS by Demitrack et al. (1991), 24h urinary free cortisol, evening plasma cortisol, cortisol responses to ACTH administration, and ACTH responses to oral CRH stimulation were all attenuated in 30 people with CFS compared to 72 healthy controls. Prior to this, Poteliakhoff (1981) demonstrated attenuated plasma cortisol levels in individuals experiencing chronic fatigue.

Relevant studies have since accumulated, and published reviews generally describe an attenuation of CSA in CFS (Cleare, 2003; Papadopoulos & Cleare, 2012; Tak et al., 2011). In a narrative review of the neuroendocrinology of CFS, Cleare (2003) questioned whether uniform HPA axis dysfunction exists in CFS, but suggested studies imply “a mild, relative hypocortisolism” in CFS, particularly using 24h urinary cortisol. This conclusion was supported in an updated review covering papers published from 2003 to 2012 (Papadopoulos & Cleare, 2012). Tak et al. (2011) presented a meta-analysis examining HPA axis activity in “functional somatic disorders” (CFS, irritable bowel syndrome, and fibromyalgia), concluding that a small but statistically significant attenuation in basal CSA (hypocortisolism) was apparent in CFS compared to healthy individuals, d = -0.14, 95% CI [-0.28, 0.00].

However, there are many studies observing no attenuation of CSA or, indeed, raised CSA in CFS (for example, Inder, Prickett, & Mulder, 2005; Papadopoulos et al., 2009; Wood, Wessely, Papadopoulos, Poon, & Checkley, 1998). Whether CSA contributes to the aetiology of CFS is currently uncertain, and which facets of basal CSA are of relevance have not been examined. It is unclear whether variations in CSA play some part in the manifestation or maintenance of chronic fatigue in other clinical populations, or fatigue experienced more generally in nonclinical populations.

5.1.2 Cortisol secretory activity

The 24h cortisol circadian rhythm has two main facets: the cortisol awakening response (CAR) and diurnal cortisol slope (DCS). The CAR describes a surge in circulating cortisol levels upon awakening, and has two components: total cortisol output, and the dynamic response usually referring to post-awakening change in cortisol level from waking level (Clow et al., 2004). Typically, total cortisol is estimated by computing the area under the curve with respect to ground (AUCg) using constituent assessments, and the dynamic response by the area under the curve with respect to increase (AUCi) (J. C. Pruessner et al., 2003) or a variant of cortisol peak minus waking level calculation. A previous review of CAR studies suggested each component may be differentially associated with various psychosocial factors (Chida & Steptoe, 2009).

The DCS models cortisol's declining trend in output throughout the rest of the day, following the CAR. An estimation of total cortisol output may also be calculated for the complete circadian rhythm (profile) using AUCg or mean cortisol levels. However, a review advised that circadian profile AUCg should predominantly be used to complement other measures of CSA as, although it provides "unique information" about average levels of cortisol, it neglects diurnal variation (Adam & Kumari, 2009, p.1431).

5.1.3 Objective of review

Examining the relationship between fatigue and CSA in real-world contexts with high levels of ecological validity is warranted given the impact fatigue has on the everyday quality of life of those who experience it (Repping-Wuts, Uitterhoeve, van Riel, & van Achterberg, 2008). Laboratory environments

can confound cortisol measurements due to the novelty stress of first-time laboratory/clinic visits and physiological reactivity induced by venipuncture in serum sampling (Schlotz, 2011). Urinary sampling offers only a summary index of cortisol output over a period of time. When incorporating strategic saliva sampling protocols, studies following research traditions such as ecological momentary assessment (Stone & Shiffman, 1994) or ambulatory assessment (Ebner-Priemer & Kubiak, 2010; Trull & Ebner-Priemer, 2013) can acquire relatively frequent repeated measures of circulating free cortisol. The present review examined studies of different CSA markers operationalized by saliva sampling in everyday life and their respective relationships with fatigue in (1) CFS; (2) other clinical populations; and (3) nonclinical populations.

5.2 Method

5.2.1 Search strategy

Systematic searches were made using the MEDLINE (Ebsco); PsycINFO (Ebsco); Embase (Ovid); Web of Science (ISI Web of Knowledge); and CINAHL (Ebsco) electronic databases (between database start and 1st June, 2012). Reference lists of relevant review articles were hand-searched. Search strings were created as follows: (1) "*fatigue*", and (2) "*saliva**" and "*cortisol*", and (3) "*circadian*" or "*diurnal*" or "*basal*" or "*daily*" or "*everyday*" or "*daytime*" or "*slope*" or "*profile*" or "*morning*" or "*awaken**" or "*evening*" or "*waking*" or "*wake*" (where * indicates truncation). Conference proceedings, dissertations, and theses were not included, although searches for published articles pertaining to these were carried out. No review protocol was published.

5.2.2 Study selection and criteria for inclusion

Articles were included if they met the following criteria: (1) adult population (≥18 years); (2) English language publication; (3) analysing original data; (4) everyday life design with salivary cortisol assessments on at least one occasion per day (fixed-occasion if only one); and (5) fatigue measured as an outcome/predictor using (i) an established self-report scale, defined as gaining at least partial inclusion in the Whitehead (2009) review of unidimensional and multidimensional fatigue measures; (ii) self-report momentary assessments; or (iii) a recognised diagnosis of CFS meeting CDC

(Fukuda et al., 1994) or Oxford (Sharpe et al., 1991) criteria. Exclusion criteria for this review were: (1) studies of cortisol reactivity to pharmacologic, physiologic, or psychosocial stimulation, unless qualifying data was provided prior to stimulation; (2) RCTs, unless qualifying baseline data; (3) inclusion of pregnant women; (4) inclusion of sleep-deprived individuals or shift workers; and (5) inclusion of participants taking steroidal medications.

5.2.3 Data extraction

After removing duplicate records, titles and abstracts were screened for obvious departures from review criteria, followed by full text screening of those remaining. This process was carried out systematically by DP, while not blinded to authors or institutions. The following data were extracted: (1) author(s); (2) year; (3) definition of fatigue; (4) study design; (5) participant characteristics (population; *n*; age; gender; inclusion/exclusion criteria); (6) number and timing of cortisol assessments; (7) number of sampling days; (8) method of maximising compliance; (9) saliva sampling procedure; (10) behavioural instructions around saliva sampling; (11) cortisol assay used; (12) fatigue measure; (13) facets of fatigue experience measured (e.g., general, physical, mental, impact); (14) variables controlled in analysis; (15) mean/*SD* of cortisol assessments and/or computations; (16) mean/*SD* of fatigue measures; (17) statistical analysis used; (18) results and conclusion. Where incomplete or unclear information was reported, attempts were made to contact corresponding authors for clarification. Where salivary cortisol data was unavailable, the study was excluded. Study screening and data extraction was completed by DP.

5.2.4 Assessment of study quality

A study quality scale was developed as no appropriate tool existed focussing on the acknowledged methodological concerns in studies using salivary cortisol assessments in everyday life (Adam & Kumari, 2009; Hansen, Garde, & Persson, 2008; Schlotz, 2011). A review of published quality tools for observational studies suggested that the three most fundamental domains for any quality tool are the appropriate selection of participants, the appropriate measurement of variables, and the appropriate control of confounding (Sanderson, Tatt, & Higgins, 2007). In addition, Sanderson et al. (2007) stated

that 86% of quality tools accounted for design-specific sources of bias, which is of particular importance here as an appropriate design must consider and account for the characteristics of the cortisol circadian rhythm. Studies should also adopt strategies to maximise compliance and utilise techniques to minimise and deal with missing assessments. The development of the quality tool was initiated with these factors in mind, with the intention of creating a valid global tool for the assessment of study quality in everyday life salivary cortisol studies.

Quality tools used in related systematic reviews were examined (Chida & Steptoe, 2009; Tak et al., 2011) and features of both were incorporated where appropriate. Three independent experts in the research area were consulted at different stages of the formulation of the scale. Each expert was emailed with a version of the scale and asked to comment on each specific item and to identify any oversights on our part.

The final items and scoring options for the 7-item quality scale are detailed in Appendix T. Scores could range from 0-16, with high scores indicating methodological quality and low scores suggesting risk of bias. For the purposes of this review, the study quality scale was applied independently by DP and WS, with discrepancies resolved by discussion.

5.2.5 Data synthesis

The heterogeneity of designs and populations between qualifying studies led to a narrative review being conducted. Three subset meta-analyses were also carried in studies with cross-sectional CFS case-control designs. These examined differences between CFS and control groups for CAR total output (meta-analysis 1); CAR dynamic response (meta-analysis 2); and circadian profile cortisol output (meta-analysis 3). It has been argued that where constituent study characteristics are similar, meta-analysis represents the most appropriate synthesis method even when the number of studies is very small (Valentine, Pigott, & Rothstein, 2010).

Standardised mean differences were calculated in all cases based on AUCg and AUCi computations, depending on the aim. Cohen's *d* was the preferred measure of standardised mean difference due to a tendency for group sizes to be quite different (McGrath, 2006). Where studies computed

AUC on more than one day, but did not compute between-day AUC mean/*SD*, the decision was made *a priori* to use only day 1 data for the meta-analysis. All cortisol values were converted to nmol/L units beforehand, and *SD*s calculated from confidence intervals where necessary. A random-effects model was considered most appropriate as it assumes varying effect sizes between studies, and permits inferences that generalise beyond those studies included (Field & Gillett, 2010). Where appropriate, Cochrane's Q and the I^2 statistic were calculated to check heterogeneity between studies. Where significant mean effect sizes were found, Fail Safe N (FSN; Rosenthal, 1979) was used to check for evidence of publication bias; FSN provides an estimated number of additional studies with null results that would be required to diminish the findings. Analyses were carried out using Review Manager Version 5.1 and bespoke meta-analysis syntax to compute FSN (Field & Gillett, 2010) with SPSS Version 20.0.

5.3 Results

5.3.1 Search and study selection

The study selection process is detailed in Figure 15. Searches revealed 514 potentially-relevant papers, and 277 duplicate records of the same report were identified and removed. Following screening and application of selection criteria, data from 19 papers were extracted. Of note, the studies by Nater et al. (2008) and Heim et al. (2009) appeared to be relevant multiple reports of the same study, testing different hypotheses. This could not be verified with the authors. The study by Nater et al. (2008) was retained as this study appeared to more closely match the purpose of this review. The study by Heim et al. (2009) was excluded.

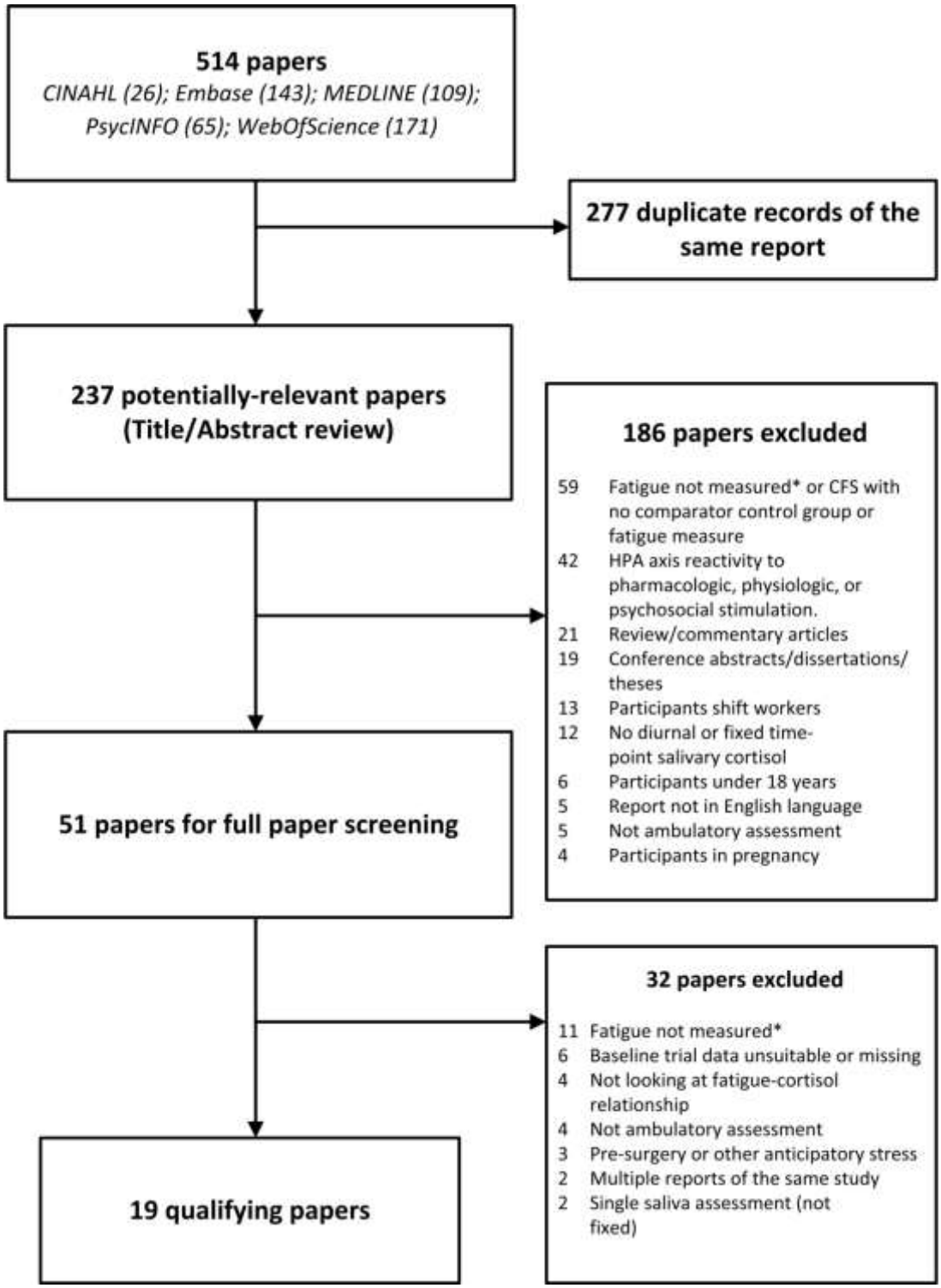


Figure 15. Flow chart of study screening and exclusion process in systematic review.

* indicates fatigue that was not operationalized by a measure meeting full or partial inclusion in the Whitehead (2009) review.

Eight studies implemented cross-sectional case-control designs in CFS (Gaab et al., 2002; Jerjes, Cleare, Wessely, Wood, & Taylor, 2005; Nater et al., 2008; Papadopoulos et al., 2009; Rahman, Burton, Galbraith, Lloyd, & Vollmer-Conna, 2011; Roberts, Wessely, Chalder, Papadopoulos, & Cleare, 2004; Strickland, Morriss, Wearden, & Deakin, 1998; A. H. Young et al., 1998), and are summarized in Table 20. Two studies qualified for meta-analysis 1 (Gaab et al., 2002; Nater et al., 2008), two for meta-analysis 2 (Nater et al., 2008; Roberts et al., 2004), and three for meta-analysis 3 (Nater et al., 2008; Papadopoulos et al., 2009; A. H. Young et al., 1998). Other markers of CSA, such as DCS, are discussed in the narrative review. Seven studies of clinical populations other than CFS qualified (Barroso, Burrage, Carlson, & Carlson, 2006; Bay & Xie, 2009; Bower et al., 2005; Dekkers, Geenen, Godaert, van Doornen, & Bijlsma, 2000; Gold et al., 2011; McLean et al., 2005; Sudhaus et al., 2009); see Table 21. Three nonclinical population studies were also included (Eek, Karlson, Garde, Hansen, & Ørbæk, 2012; Kumari et al., 2009; Lindeberg et al., 2008); see Table 22. For clarity, all saliva samples provided upon awakening are henceforth referred to as T0, T30 (30 minutes after awakening), and so on.

5.3.2 Study quality

Quality Scale item scores for each study are presented in Appendix U and the total score in the final column of Tables 20-22. Respective study populations and salivary cortisol assessments were generally well-defined and reported within the studies reviewed. Several studies failed to incorporate any method towards maximizing sampling adherence, and some potentially confounded data by overlooking actual waking time when requesting early-morning samples. One potential source of bias apparent in the majority of studies reviewed was the omission of any analytical plan for dealing with missing assessments. The median score for the Quality Scale was 7 out of 16, with scores for CFS studies ranging from 3 to 11, and for other populations from 4 to 14.

Table 20 Characteristics of Chronic Fatigue Syndrome Case-Control and Cross-Sectional Studies in Systematic Review

Study	Design	Sample: N (% female), mean age (*SD*)	Psychiatric comorbidity	Saliva sampling protocol	Basal cortisol computations	Main findings	Quality score[a]
Gaab et al. (2002)	Case-control, cross-sectional	*CFS*: 21 (52), 36.0 years (4.5); *Control*: 21 (52), 35.2 years (4.5).	1 CFS had current episode of MDD; 7 CFS history of MDD; 4 CFS history of anxiety disorder.	Days: 2 Assessments: 9 (T0, T15, T30, T45, T60, 0800h, 1100h, 1500h, 2000h).	CAR (AUCg); Circadian profile (AUCg).	No difference between groups on any basal cortisol measure. Graphical representation seemed to suggest flatter CARs in CFS than controls, although not statistically tested.	8
Jerjes et al. (2005)	Case-control, cross-sectional	*CFS*: 15 (53), 35 years (7.9); *Control*: 20 (50), 33 years (11.3)	Individuals with current major depression or anxiety disorders excluded.	Days: 1 Assessments: 6 (T0 (0600h), 0900h, 1200h, 1500h, 1800h, 2100h).	Circadian profile ("goodness of fit" to cosinor curve, midline estimate statistic of rhythm (MESOR), amplitude, acrophase); Fixed Time points.	Evidence for reduced CSA in CFS compared to controls in terms of both MESOR and amplitude (half the cortisol peak minus nadir). Lower cortisol levels observed at all time-points, except 2100h.	6
Nater et al. (2008)	Case-control, cross-sectional	*CFS*: 75 (77), 43.9 years (*SD* not given); *Control*: 110 (75), 44.8 years (*SD* not given).	21.3% of CFS group fulfilled diagnostic criteria for MDD.	Days: 1 Assessments: 3 (T0, T30, T60).	CAR (AUCg, AUCi, slope, peak-waking)	Evidence that CFS may be associated with attenuated CAR, but effect only present in women. AUCi, slope, and peak minus waking levels all attenuated in female participants.	11

Study	Design	Sample: N (% female), mean age (*SD*)	Psychiatric comorbidity	Saliva sampling protocol	Basal cortisol computations	Main findings	Quality score[a]
Papadopoulos et al. (2009)	Case-control, cross-sectional	*CFS*: 18 (56), 39.1 years (8.2); *Control*: 20 (65), 39.5 years (11.4)	9 CFS with comorbid depression.	Days: 1[b] Assessments: 4 (0800h, 1200h, 1600h, 2000h).	Circadian profile (AUCg)	AUCg significantly higher in CFS compared to controls (CFS with and without comorbid MDD both higher AUCg than controls).	6
Rahman et al. (2011)	Case-control, cross-sectional	*CFS*: 15 (87), 32.5 years (11.1); *Control*: 15 (67), 35.6 years (13.9)	MDD and psychosis excluded.	Days: 1 Assessments: 4 (T30 (to have occurred 0600h-0800h), 1200h, 1800h, 2200h).	Fixed time-points only	No difference at any time-point.	11
Roberts et al. (2004)	Case-control, cross-sectional	*CFS*: 56 (63), 39.4 years (11.0); *Control*: 35 (60), 34.9 years (12.8)	22 CFS with comorbid depression	Days: 1 Assessments: 5 (T0, T10, T20, T30, T60).	CAR (AUCi)	Attenuated CAR in CFS group compared to controls.	9
Strickland et al. (1998)	Case-control, cross-sectional	*CFS*: 14 (100), 36 years (11); *MDD*: 26 (100), 34 (6); *Control*: 131 (100), 34 years (7)	10 CFS mild or moderate depressive episodes.	Days: 2 Assessments: 2 (1100h, 2100h)	Fixed time-points (morning, evening)	No difference in 1100h cortisol between groups (trend for lower cortisol in CFS). CFS without depression significantly lower cortisol at 1100h than controls. Cortisol lower for CFS at 2100h than both controls and depressed group.	8

Study	Design	Sample: N (% female), mean age (*SD*)	Psychiatric comorbidity	Saliva sampling protocol	Basal cortisol computations	Main findings	Quality score[a]
Torres-Harding et al. (2008)	Cross-sectional	*CFS*: 108 (83), Men = 39.6 (11.5); Women = 43.9 (11.4)	Not stated, but individuals with psychiatric disorder and taking antidepressants clearly present within sample.	Days: 1 Assessments: 5 (T0, T45, 0900h, 1600h, 2100h).	Circadian profile (Mean day cortisol, slope, physician classification of profile as "normal" or "abnormal").	No relationship between level of fatigue within CFS and mean cortisol or DCS. Those with "abnormal" cortisol profiles had higher fatigue severity than those with "normal" profiles.	3
A. H. Young et al. (1998)	Case-control, cross-sectional	*CFS*: 22 (45), 39 years (8.8); *Control*: 22 (45), 38 years (8.0)	Current depressive or anxiety disorder excluded. 2 CFS definite history of MDD, 10 CFS probable history of MDD.	Days: 1 Assessments: 4 (0800h, 1200h, 1600h, 2000h).	Circadian profile (AUCg)	No differences between groups.	6

Note. MDD represents major depressive disorder; T*x*, *x* minutes after awakening. CAR, cortisol awakening response; DCS, diurnal cortisol slope; AUCg, area under the curve ground; AUCi, area under the curve increase; CSA, cortisol secretory activity.

[a] See Quality Scale items and respective scores in Appendices T and U for further information. Maximum score is 16.

[b] Additional measurement day in study but was post-dexamethasone.

Table 21 Characteristics of Clinical Population Studies in Systematic Review

Study	Design	Sample population, N (% female), mean age (*SD*)[a]	Saliva sampling protocol	Basal cortisol computations	Fatigue measure	Main findings	Quality score[b]
Barroso et al. (2006)	Cross-sectional	*HIV-positive,* 40 (28), 39.45 years (*SD* not given)	Days: 1 Assessments: 3 (0700h, 1500h, 2200h).	None. Only visual DCS trends.	HIV-related fatigue scale	An upwards cortisol slope over the day (*n* = 5) had highest levels of fatigue (not empirically tested).	4
Bay and Xie (2009)	Cross-sectional	*Mild-to-moderate traumatic brain injury,* 75 (48), Age not reported.	Days: 1 Assessments: 4 (0800h, 1200h, 1600h, 2200h).	Circadian profile (AUCg).	Fatigue subscale of the Profile of Mood States.	AUCg did not predict concurrent fatigue.	6
Bower et al. (2005)	Cross-sectional	*Breast cancer survivors,* 29 (100), Fatigued = 58.2 years; Not fatigued = 61.8 years (*SD* not given)	Days: 2 Assessments: 4; (T0 (not while still in bed), 1200h, 1700h, 2200h).	Circadian profile (AUCg), mean cortisol across days, DCS.	Fatigue subscale of RAND SF-36.	Higher levels of fatigue were assoc. with flatter cortisol slopes (consequence of reduced cortisol decline in the evening).	8
Dekkers et al. (2000)	Cross-sectional	*Rheumatoid arthritis (recently diagnosed),* 25 (76), 55.2 years (*SD* not given)	Days: 2 Assessments: 9 (T0, T15, T30, T45, 1000h, 1200h, 1430h, 1700h, 1930h).	CAR (slope); Circadian profile (AUCg, intra-individual standard deviation).	Momentary fatigue ("I feel tired"; 1-5).	Steeper cortisol increases post-awakening associated with greater fatigue. Negative correlation between T0 cortisol and fatigue at every time-point later in the day.	7

Study	Design	Sample population, N (% female), mean age (*SD*) [a]	Saliva sampling protocol	Basal cortisol computations	Fatigue measure	Main findings	Quality score [b]
Gold et al. (2011)	Cross-sectional	*Multiple sclerosis (relapsing-remitting),* 44 (100), 35.8 years (0.7)	Days: 2 [c] Assessments: 9 (T0, T15, T30, T45, T60, 1100h, 1500h, 2000h, 2200h).	CAR (AUCg); Circadian profile (AUCg).	Wurzburg Fatigue Inventory for MS	No evidence of a relationship between CSA and fatigue.	7
McLean et al. (2005)	Cross-sectional	*Fibromyalgia,* 16 (100), 43 years (9)	Days: 2 Assessments: 5 (T0, T60, 5hrs after waking, late afternoon (1500h-1600h), 30mins before bed).	Fixed time points only	Momentary fatigue (wording unavailable; 1-100).	No evidence of a relationship between momentary fatigue and concurrent levels of cortisol at any time-point.	14
Sudhaus et al. (2009)	Cross-sectional	*Acute and chronic lower back pain (ALBP; CLBP),* Two groups: ALBP = 19 (63), 39.8 years (12.3); CLBP = 24 (71) 38.3 years (11.4)	Days: 2 Assessments: 5 (T0, T15, T30, T45, T60).	CAR (AUCg)	General Fatigue Scale of the Multi-dimensional Fatigue Inventory	CAR AUCg positively correlated with fatigue in ALBP, but negative correlation in CLBP. Within the CLBP group, highly fatigued individuals had attenuated CAR compared to those with low fatigue.	10

Note. CAR represents cortisol awakening response; DCS, diurnal cortisol slope; AUCg, area under the curve ground; AUCi, area under the curve increase; T*x*, *x* minutes after awakening.

[a] *N*, sex, age, and cortisol computations reported or used in relevant analyses (patient group in all cases).

[b] See Quality Scale items and respective scores in Appendices T and U for further information. Maximum score is 16.

[c] A third day was post-dexamethasone.

Table 22 Characteristics of General Population Studies in Systematic Review

Study	Design	Sample population, N (% female); mean age (*SD*)	Saliva sampling protocol	Basal cortisol computations	Fatigue measure	Main findings	Quality score[a] (16)
Eek et al. (2012)	Cross-sectional	*Workers,* 581 (61); 46.3 years (10.7)	Days: 1 Assessments: 3 (T0, T30, 2100h)	CAR (morning peak, % increase between T0 and T30, mean of T0 and T30); Circadian profile (mean cortisol, evening minus morning peak) Fixed time-points	Swedish Occupational Fatigue Inventory	% increase in CAR positively associated with lack of energy, lack of motivation, and physical exertion SOFI-20subscales. Physical exertion negatively associated with T0 cortisol. Associations stronger in women than in men (no statistically significant results in male-only analyses).	5
Kumari et al. (2009)	Cross-sectional and longitudinal: (1) Phase 6 (P6; 2001); (2) Phase 7 (P7; 2003-2004); (3) Phase 8 (P8; 2006)	*Whitehall II Cohort*, 4299, "Fatigued" (22.3% female); 61.4 years (*SD* not given); "Not fatigued" (34.2% female); 59.8 years (*SD* not given)	Days: 1 Assessments: 6 (T0, T30, T2.5hrs, T8hrs, T12hrs, "bedtime"). Measured at P7 only.	CAR (T30 minus T0); Circadian profile (DCS, using all samples except T30); Fixed time-points.	Vitality subscale of SF-36 (cut-off ≥50). Fatigue measured at All Phases.	Longitudinal (P6/P7): P6 fatigue not predictive of P7 cortisol. Cross-sectional (P7): CAR similar in fatigued and non-fatigued groups. T0, T30, and DCS all lower or flatter in fatigued group (T0 cortisol association accounted for by other health measures). "Bedtime" cortisol higher in fatigued group. Longitudinal (P7/P8): Flatter P7 DCS associated with persistent fatigue P7-P8 and with new fatigue at P8.	9

Study	Design	Sample population, N (% female); mean age (*SD*)	Saliva sampling protocol	Basal cortisol computations	Fatigue measure	Main findings	Quality score[a] (16)
Lindeberg et al. (2008)	Cross-sectional	*Workers*, 78 (73); "Exhausted" 45 years (8); "Not exhausted" 46 years (9).	Days: 1 Assessments: 3 (T0, T30, 2100h).	Circadian profile (peak minus nadir, mean cortisol); Fixed time-point.	Vitality subscale of SF-36[b] (cut-off ≥16)	Statistically significant negative correlation between vitality score and peak-nadir cortisol. Peak-nadir lower in fatigued group. No difference in mean cortisol or cortisol at each time-point between groups.	7

Note. T*x* indicates samples provided *x* minutes after awakening; CAR represents cortisol awakening response; DCS, diurnal cortisol slope,

[a] See Quality Scale items and respective scores in Appendices T and U for further information. Maximum score is 16.

[b] inverted SF-36 (fatigue measured in positive direction, such that high scores indicate high fatigue).

5.3.3 Chronic fatigue syndrome

Eight case-control studies were selected, incorporating 636 participants. All studies where fatigue severity was measured revealed statistically significant differences in fatigue scores between groups. One additional qualifying study did not incorporate a control group and used a fatigue measure to distinguish levels of fatigue within their CFS population (Torres-Harding et al., 2008). Comorbid psychiatric disorders were present within the CFS group in six out of the nine studies included.

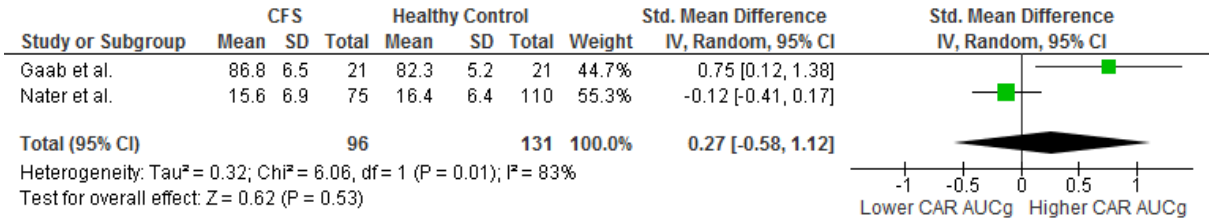
Three studies assessed the CAR (Gaab et al., 2002; Nater et al., 2008; Roberts et al., 2004). In meta-analysis 1, across both studies (Gaab et al., 2002; Nater et al., 2008) (CFS n = 96, healthy control n = 131) the mean between-group effect size was d = 0.27, 95% CI [-0.58, 1.12], with no significant overall effect (Z = 0.62, p =.53). Figure 16(i) reflects this, showing that, although the study by Gaab et al. (2002) showed a higher CAR AUCg in CFS, this was not replicated in the study by Nater et al. (2008). However, meta-analysis 1 likely reflected a heterogeneous sample of studies (I^2 = 83%; Cochran's Q χ^2 = 6.06, p = .01).

In meta-analysis 2, across both studies (Nater et al., 2008; Roberts et al., 2004) (CFS n = 131, healthy control n = 145) the mean between-group effect size was d = -0.34, 95% CI [-0.58, -0.09], and revealed a significant overall effect (Z = 2.72, p = .006, FSN = 4). Both studies individually found a significant between-group difference for AUCi (see Figure 16(ii)) reflecting an attenuated cortisol response to awakening in CFS from T0. Meta-analysis 2 did not reflect heterogeneity (I^2 = 0%; Cochran's Q χ^2= 0.21, p = .65).

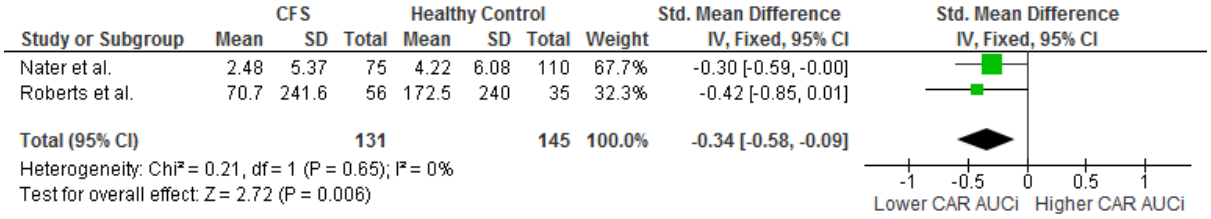
Four studies examined group differences in the circadian cortisol profile (Gaab et al., 2002; Jerjes et al., 2005; Papadopoulos et al., 2009; A. H. Young et al., 1998). Three of these (Gaab et al., 2002; Papadopoulos et al., 2009; A. H. Young et al., 1998) estimated total cortisol output (AUCg) and were entered into meta-analysis 3. The remaining paper (Jerjes et al., 2005) carried out cosinor rhythm analysis. Across all three studies (CFS n = 61, healthy control n = 63) the mean between-group effect size was d = 3.18, 95% CI [0.38, 5.98], reflecting a significant overall effect (Z = 2.23, p =.03). Figure 16(iii) depicts this. Heterogeneity was high (I^2 = 97%; Cochran's Q χ^2 = 64.12, p < .001), and

one study (Papadopoulos et al., 2009) provided outlying results. The study by Papadopoulos et al. (2009) was excluded in a subsequent meta-analysis (Figure 16(iv)) where heterogeneity between studies was rejected (I^2 = 0%; Cochran's Q χ^2 = 0.54, p = .46). The remaining studies (CFS n = 43, healthy controls n = 43) reflected a mean effect size of d = 0.01, 95% CI [-0.43, 0.42], with no significant overall effect (Z = 0.03, p = .98).

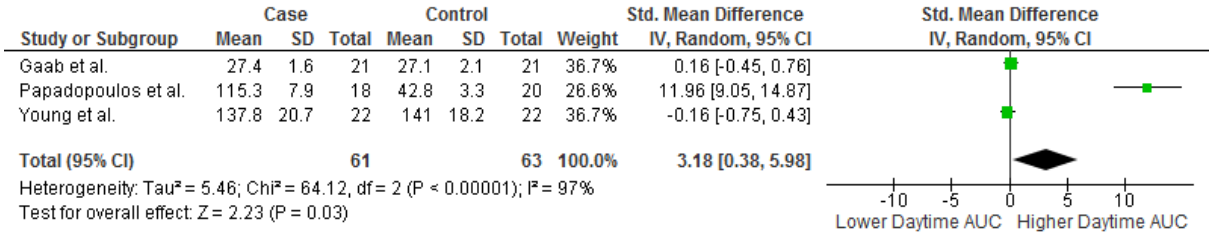
(i)



(ii)



(iii)



(iv)

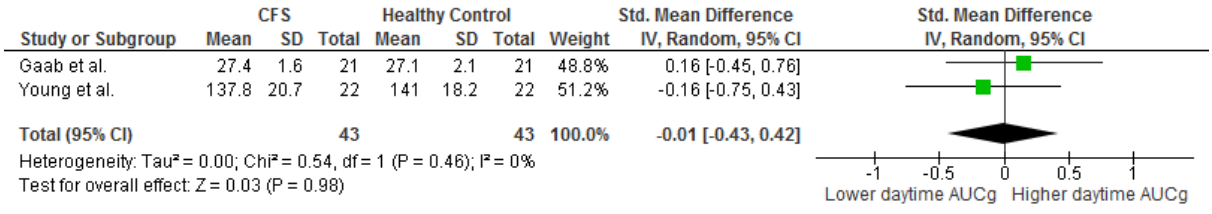


Figure 16. Forest plots for (i) CAR estimated total cortisol output; (ii) CAR estimated dynamic response; and circadian profile total cortisol output in CFS (iii) with, and (iv) without Papadopoulos et al. (2009) paper.

Only one of the CFS case-control studies examined cortisol variability within the whole day, using cosinor analysis to perform individual and population mean computations as detailed in the relevant column of Table 20 (Jerjes et al., 2005). Midline estimate statistic of rhythm (MESOR), defined as the rhythm-adjusted mean cortisol, was shown to be significantly lower in the CFS group than the control group. Amplitude, which was defined as half the difference between daytime cortisol peak and nadir, was also lower in CFS than healthy controls.

Examination of fixed time-based group comparisons in the case-control studies were mixed, with one study revealing lower CSA in the CFS group at every assessment (T0, 0900h, 1200h, 1500h, 1800h) except 2100h, where there was no difference (Jerjes et al., 2005). Another study observed lower levels of CSA in the CFS group only at 2100h (and not at 1100h: the only other assessment) (Strickland et al., 1998). In the remaining study performing such comparisons, no statistically significant differences were observed between groups at any assessment (T30, 1200h, 1800h, 2200h) (Rahman et al., 2011).

One CFS study without a control group measured levels of fatigue within the sample (Torres-Harding et al., 2008). The study found no relationship between fatigue and either the daily mean cortisol computation or the DCS. However, using a classification approach, a physician categorized participant circadian cortisol profiles as "normal" or "abnormal" (based on "expected" ranges of cortisol profiles), and found that "abnormal" profiles were associated with higher levels of fatigue severity. "Abnormal" profiles did not reflect a specific type of cortisol growth trend but did contain at least one of the following features: (1) "cortisol patterns or results exhibiting divergent peak times"; (2) "decreases in cortisol level followed by sudden increases"; or (3) "general attenuation of cortisol pattern" (Torres-Harding et al., 2008, p.166). The validity of this classification is unknown, and has not been replicated elsewhere.

5.3.4 Fatigue in other clinical groups

Table 21 shows the characteristics of each of the studies incorporating clinical populations. Previous evidence has suggested a high prevalence of fatigue in each of the conditions represented here: MS (Lerdal et al., 2003); rheumatoid arthritis (Belza, 1995); fibromyalgia (White, Speechley, Harth, &

Ostbye, 2000); chronic lower back pain (Fishbain et al., 2004); traumatic brain injury (Oullet & Morin, 2006); and HIV (Breitbart, McDonald, Rosenfeld, Monkman, & Passik, 1998). Most studies recruited healthy control groups, but the relationship between CSA and fatigue within the control group was not reported in any study.

The CAR was examined in three of the seven studies (Dekkers et al., 2000; Gold et al., 2011; Sudhaus et al., 2009). All of these studies computed CAR AUCg, while one also operationalized individual CAR dynamic responses as regression slope parameters (Dekkers et al., 2000). Only one study found a relationship between fatigue level and AUCg (Sudhaus et al., 2009). In this study, the chronic lower back pain group had associations between greater fatigue and attenuated CAR output, but the acute lower back pain group showed the opposite: greater fatigue was associated with elevated CAR output. A median split on the Multidimensional Fatigue Inventory (MFI; Smets et al., 1995) showed there were CSA differences between high and low fatigue only within the chronic pain group, with an attenuated CAR AUCg in those categorised as high fatigue compared to low fatigue. In the remaining two studies, CAR AUCg was not associated with fatigue level (Dekkers et al., 2000; Gold et al., 2011).

As mentioned, the CAR dynamic response was assessed in only one study (people with rheumatoid arthritis). Higher slope parameters (steeper cortisol responses to awakening) were associated with greater levels of daytime fatigue, but not daytime fatigue variability (Dekkers et al., 2000). Momentary fatigue assessments were incorporated in this study to compute these variables. The authors also performed time-lagged associations between fixed momentary assessments, finding correlations of T0 cortisol with fatigue at any assessment during the day at around $r = -.35$, with successive cortisol assessments up until 1000h becoming less negatively correlated with fatigue. These time-lagged analyses imply that increased CSA within the first hour of awakening, particularly precisely upon awakening, may be related to how fatigue is experienced throughout the rest of the day in rheumatoid arthritis.

Five of the seven studies examined at least one facet of the circadian cortisol profile (Bay & Xie, 2009; Bower et al., 2005; Dekkers et al., 2000; Gold et al., 2011; McLean et al., 2005) and one study characterised circadian cortisol profiles into different “cortisol trends” using visual categorisations after having

problems having saliva samples from their HIV population analysed (Barroso et al., 2006). No study found any relationship between total estimated cortisol output over the day and fatigue. Bower et al. (2005) found no differences in cortisol AUCg or mean cortisol between those categorised as fatigued (score < 50) or not fatigued (score >70) by the energy/fatigue subscale of the RAND SF-36 (Hays, Sherbourne, & Mazel, 1993). Gold et al. (2011) found that cortisol AUCg was not a significant predictor of cross-sectional fatigue. Bay and Xie also observed that cortisol AUCg did not make a significant contribution to a statistical model of fatigue.

In terms of CSA daily variability, Bower et al. (2005) reported that higher levels of fatigue severity were associated with a flatter DCS. Those within the fatigued group had a flatter DCS than those who were not fatigued, with approximately 25% of between-subject DCS variation accounted for by fatigue group. Dekkers et al. (2000) found that the intra-individual standard deviation of cortisol ("daytime cortisol variability") was not associated with level of fatigue.

The only study in an HIV population had participants' saliva samples refused by the planned analysis laboratory (Barroso et al., 2006). The alternative laboratory required samples be diluted 16 times, which led the authors to render only primitive analyses appropriate. Cortisol profiles were categorised into different trends: (1) "normal downward trend"; (2) "afternoon fall"; (3) "afternoon peak"; (4) "upward trend". Although no statistical testing was conducted between groups, it was apparent that those with a "normal downward trend" had the lowest ratings of fatigue on the Fatigue Severity Index, computed from the intensity and consequences subscale of the HIV-Related Fatigue Scale (Barroso & Lynn, 2002).

Two studies looked at associations between fixed-occasion salivary cortisol assessments and momentary fatigue (Dekkers et al., 2000; McLean et al., 2005). Neither showed any associations between cortisol and concurrent fatigue, although lagged associations, as detailed previously, were apparent in the rheumatoid arthritis study (Dekkers et al., 2000). Bower et al. (2005) found that between "fatigued" and "not fatigued" participants, there were differences in salivary cortisol levels only in the evening, with cortisol levels marginally higher in the fatigued group at 1700h and reached significance at 2200h.

5.3.5 Fatigue in nonclinical populations

Three studies examined the relationship between fatigue and CSA in nonclinical populations (Eek et al., 2012; Kumari et al., 2009; Lindeberg et al., 2008). Table 22 depicts study characteristics for each. Two studies examined facets of the CAR in relation to fatigue (Eek et al., 2012; Kumari et al., 2009). Eek et al. (2012) computed mean cortisol between T0 and T30 as an estimation of CAR output, and computed a percentage increase from T0 to T30 as a marker of the dynamic cortisol response to waking. The study found no correlation between CAR output and fatigue on five subscales (lack of energy; physical exertion; physical discomfort; lack of motivation; sleepiness) of the Swedish Occupational Fatigue Inventory (SOFI; Åhsberg, Garnberale, & Kjellberg, 1997). However, positive associations between the CAR dynamic response and three of the SOFI subscales (lack of energy; physical exertion; lack of motivation) were found, but in subsequent analyses these were found to be present in female participants only. Conversely, Kumari et al. (2009) operationalized the CAR dynamic response by calculating peak (T30) minus nadir (T0), and categorised participants into two groups based on a cut-off of 50 on the vitality subscale of the SF-36 (Ware & Sherbourne, 1992). No differences were found for CAR peak minus nadir between the fatigued and not fatigued groups.

In the third study, salivary assessments were made at T0, T30, and 2100h (Lindeberg et al., 2008). Participants were split into two groups based on a cut-off of 16 on the vitality subscale of the SF-36. There were no differences in daily mean CSA between groups, but the more fatigued group did have significantly lower peak minus nadir values (flatter cortisol profiles). In addition, the study found a negative correlation between vitality score and peak minus nadir values such that they became smaller, and cortisol profiles flatter, with increasing fatigue (vitality scores were inverted such that higher scores indicated more fatigue). Eek et al. (2012) did not find any associations between fatigue and mean cortisol level or peak minus nadir cortisol values. Kumari et al. (2009) also examined the circadian rhythm, but in terms of DCS, finding that slopes were flatter in the more fatigued group.

All three studies examined associations between fatigue and salivary cortisol at fixed assessment time points (Eek et al., 2012; Kumari et al., 2009; Lindeberg et al., 2008). Kumari et al. (2009) observed lower cortisol levels in

the more fatigued group at T0 and T30, and higher cortisol levels at bedtime, with no differences for any of the fixed afternoon assessments (T2.5hrs; T8hrs; T12hrs). Eek et al. (2012) found a negative correlation between the physical exertion subscale of the SOFI scale and cortisol at T0 (in females only), but no associations with any other subscale or at any other time point. No differences were found between groups at any occasion (T0, T30, nor 2100h) in the study by Lindeberg et al. (2008).

Only one study included in this review had a longitudinal facet to its design (Kumari et al., 2009), and was a large epidemiological study that recruited from the Whitehall II cohort (see Marmot & Brunner, 2005) categorizing each individual as fatigued or not fatigued, as already outlined, at Phase 6 (2001), Phase 7 (2003/04), and Phase 8 (2006). Salivary cortisol assessments were made only at Phase 7. The study revealed no association between Phase 6 levels of fatigue and Phase 7 CAR peak minus nadir or slope parameter. As has been stated, cross-sectional associations at Phase 7 were apparent, but the between-group differences for T0 cortisol disappeared once health measures were included in the model. A flatter DCS at Phase 7 was associated with persistent fatigue through Phases 7 and 8, as well as predictive of the onset of new fatigue at Phase 8.

5.4 Discussion

In this review, 19 studies published before June 2012 were included examining the relationship between fatigue experience and different markers of unstimulated (basal) CSA in everyday life, using salivary cortisol ambulatory assessments. An attenuation of cortisol diurnal variability was found to be important in relation to fatigue in most constituent studies. Measures of total cortisol output, which discard information about diurnal variability (Adam & Kumari, 2009), were rarely associated with fatigue in those studies reviewed. These findings were supported by three meta-analyses of CFS case-control studies and by a narrative synthesis of clinical and nonclinical studies; although conclusions were less clear in the non-CFS clinical groups.

5.4.1 Chronic fatigue syndrome

Evidence from eight case-control studies and one case-only study indicated decreased within-person CAR and circadian cortisol variation within

CFS. An attenuation of diurnal variability concurs with the most recent review of endocrine dysfunction in CFS (Papadopoulos & Cleare, 2012). However, the present review differs in suggesting that CSA markers that neglect cortisol variability, such AUCg, have limited relevance to CFS; published reviews have thus far proposed at least modest reductions in total cortisol output in CFS (Cleare, 2003; Papadopoulos & Cleare, 2012; Tak et al., 2011).

A lack of qualifying studies prevents overly-robust conclusions being drawn but, taken together with the narrative synthesis, these observations would seem consistent and could inform the sorts of cortisol markers computed in future empirical studies in CFS. The research area requires more studies that incorporate prudent steps to maximising participant adherence to study designs to fully validate these findings, as this is a particular weakness in those studies reviewed here. Consensus on how best to operationalize CSA has not yet been reached. Studies continue to report various models of cortisol activity from constituent assessments, as was demonstrated by two of the nonclinical studies which had the same one-day salivary cortisol assessment design (T0; T30; 2100h). Despite both wanting to test associations between fatigue and cortisol levels, each computed different markers of CSA and came to opposing conclusions (Eek et al., 2012; Lindeberg et al., 2008). The heterogeneity of CSA measurement is not a criticism exclusive to CFS research, and is apparent in other populations and, indeed, outside of the sphere of fatigue research.

Childhood trauma may contribute to the aetiology of CFS. One study has identified childhood maltreatment as a risk factor for CFS, and this relationship may be mediated by HPA axis activity (Heim et al., 2009). Indeed, Heim et al. (2009) found only those with CFS who had experienced childhood trauma had an attenuated CAR compared to controls.

Finally, latent class analysis performed in a study of individuals with CFS has identified up to five subgroups in CFS, indicating that CFS is likely to be an umbrella term (Cella, Chalder, & White, 2011). Earlier work by Jason et al. (2005) argued that not employing subtypes in CFS research may be contributing to general inconsistencies in findings within this population. Future studies of salivary cortisol in CFS may benefit from performing subgroup analyses based upon variables such as depression, sleep quality, gender, and pain sensitivity.

5.4.2 Other clinical and nonclinical groups

AUCg measures were associated with fatigue in only the study of CSA in lower back pain (Sudhaus et al., 2009), finding associations in alternate directions depending on having chronic (negative association) or acute (positive association) lower back pain (Sudhaus et al., 2009). The authors speculated upon reasons for this differential association: (1) when fatigue accompanies pain, daily activities become more stressful and there is increased CSA in acute pain via the HPA axis stress response; and (2) when this stress evolves to become more chronic, this reflects in the HPA axis dysfunction (hypocortisolism) that was observed in the chronic lower back pain group.

Contrary to the apparent trend in CFS studies, two studies found larger increases upon awakening were associated with greater fatigue (Dekkers et al., 2000; Eek et al., 2012). In rheumatoid arthritis, Dekkers et al. (2000) argued that this potentially reflected an "adaptive response of the HPA system" to the fatigue experienced by elevating morning cortisol levels to facilitate higher energy availability (p. 367). Eek et al. (2012) found a similar relationship with fatigue in a nonclinical population, but secondary analyses revealed this to exist only in women. No association between the awakening response and fatigue level was found in a similar analysis with a different nonclinical population (Kumari et al., 2009).

There was evidence of reduced CSA diurnal variability in fatigued individuals in cross-sectional analyses with nonclinical populations. The one longitudinal study included indicated that fatigue severity did not predict CSA 2-3 years later, but some facets of CSA (flattened DCS and low waking cortisol) were predictive of concurrent and persistent or new fatigue 2-3 years later (Kumari et al., 2009). These findings suggests that changes in basal CSA may occur prior to the onset of (or early in the process of developing) fatigue; a position which would appear at odds with the argument made in the rheumatoid arthritis study that the HPA axis may become more active in order to facilitate increased energy metabolism when experiencing fatigue (Dekkers et al., 2000).

5.4.3 Relevant salivary cortisol markers in fatigue research

In relation to fatigue experience, the relative importance of within-day cortisol variability compared to estimations of total cortisol output was apparent across all populations. A significant association between fatigue and total cortisol output was only apparent within an outlying study in CFS (Papadopoulos et al., 2009), suggesting that this facet of CSA is not relevant to everyday fatigue experience: chronic or otherwise. In a review of psychosocial factors and the CAR, it was determined that AUCi, rather than AUCg, likely represents a "more appropriate measure for assessing HPA activation following waking in relation to psychosocial factors" (Chida & Steptoe, 2009, p. 275). Further, estimates of total cortisol output may be less important than measures of circadian rhythm or variability as outcomes as these latter types may be prominent indicators of regulatory competence (Sephton & Spiegel, 2003). Taken together, this suggests future studies in this field should observe at least one marker of CSA variability, rather than exclusively make comparisons of fixed occasion cortisol levels or estimates of total cortisol output.

In terms of causality, it was not possible to draw firm conclusions due to the lack of longitudinal ambulatory studies that have been conducted thus far. It would seem from the one study available that a flattened DCS may predict future fatigue (Kumari et al., 2009), which complements evidence that cortisol replacement therapy can reduce short-term fatigue in CFS (Cleare et al., 1999; McKenzie et al., 1998). However, there have been suggestions in previous reviews that HPA axis activity (or dysfunction) may not be at the core of CFS, but instead occur as a result of certain behavioural changes associated with the illness (Cleare, 2004). Inferences from two studies (Dekkers et al., 2000; Sudhaus et al., 2009), where CSA variability was actually increased in relation to fatigue, implied that fatigue may precipitate an adaptive or maladaptive alteration in HPA axis activity. It is not yet possible to discern with any certainty whether specific characteristics of CSA may trigger fatigue experience, or vice versa.

5.4.4 Limitations and future research

There are several limitations to acknowledge in this systematic review. A lack of published studies, incomplete reporting, and heterogeneous CSA

measures all contributed to the small number of studies included in each meta-analysis. Each study also contained relatively few participants so, for both these reasons, interpretations should be taken with caution. It is also important to note that the bidirectional relationship between the immune system and the HPA axis was not considered within this review; it is likely that variations in immune activity will be relevant to the experience of fatigue in all populations, whether that be directly or indirectly via changes in HPA axis function (Bansal, Bradley, Bishop, Kiani-Alikhan, & Ford, 2012).

There are many different markers of CSA that can be computed from constituent assessments and reported, and despite some efforts to suggest reporting consensus (for example, Clow et al., 2004) the variety of approaches to analysis create difficulties in making between-study comparisons in reviews. Some studies chose to examine only fixed-occasion assessments rather than compute other, potentially more informative, markers of CSA such as variants of the AUC or DCS. Reaching consensus in the way researchers compute and report cortisol outcomes should be a priority for the research area (Clow et al., 2010).

Psychiatric comorbidity was also present within some study samples, which is particularly important within CFS where a recent study has reported comorbidities of depression and/or anxiety disorders in around half of CFS patients (Cella, White, Sharpe, & Chalder, 2013), and previous reports of psychiatric comorbidities have been as high as 75% (Wessely, Chalder, Hirsch, Wallace, & Wright, 1996). Salivary cortisol may be secreted at greater levels in those with depression (Knorr et al., 2010), and Papadopoulos et al. (2009) felt latent mild or moderate depression with their CFS sample may have contributed towards their unexpected outlying results. Future research should focus on longitudinal designs, where possible, as there is a paucity of this type of evidence available and may contribute to understanding the direction of causality between fatigue and CSA.

Future research may also consider examining whether CSA is associated with fatigue within-subjects. For example, a study with 156 older adults (M = 57 ±4.5 years) demonstrated attenuated morning cortisol level (single sample upon awakening) was associated with greater levels of "physical symptoms/tiredness" later in the same day (Adam, Hawkley, Kudielka, & Cacioppo, 2006). This study did not qualify for the systematic review due to

the measure incorporating physical symptoms (headache, stomach ache, cough), but these results suggest that CSA in the early-morning may influence fatigue experienced later in the day. Whether this association is present in clinical populations with fatigue measured explicitly may be an informative future direction for research.

5.4.5 Conclusion

The systematic review suggests that CSA variability (dynamic responses to waking and variations in circadian activity) may be the most relevant facet of CSA within fatigued individuals across all populations. It appears that attenuation in this variability is associated with CFS. The review was unable to unequivocally support the hypocortisolism hypothesis in CFS, although it would seem that smaller increases in cortisol after awakening are apparent within this clinical group. Longitudinal evidence was minimal, so no robust conclusions could be drawn regarding causality underlying the association between unstimulated CSA and fatigue.

Chapter 6: Salivary cortisol and fatigue in relapsing-remitting MS

Chapter 5 suggested any relationship between fatigue and salivary cortisol measures across populations is likely with attenuated diurnal cortisol variability rather than with total cortisol output. Chapter 6 presents an examination of the relationship between cortisol secretory activity (CSA) in everyday life and fatigue at baseline and 6-month follow-up in pwRRMS and healthy individuals.

6.1 Introduction

Few studies have examined the relationship between MS fatigue and activity of the HPA axis. Cortisol is relevant to fatigue as it promotes energy metabolism (Khani & Tayek, 2001; Sapolsky et al., 2000) and low-dose corticosteroids have been shown to reduce fatigue in chronic fatigue syndrome (Cleare et al., 1999; McKenzie et al., 1998). Although an attenuation of cortisol diurnal variability seemed most relevant to fatigue across populations in the systematic review (Chapter 5), there was no association of cortisol output or diurnal cortisol slope (DCS) with fatigue in the one qualifying study in MS (Gold et al., 2011). There are no other studies of basal CSA (urinary, serum, or saliva measurement) involvement in MS fatigue.

Gold et al. (2011) recruited 44 remission-phase RRMS participants, 10 of whom had major depressive disorder (MDD) comorbidity established by clinical psychologist using the Structured Clinical Interview for DSM-IV Axis I Disorders (First, Spitzer, Gibbon, & Williams, 1997). Salivary cortisol assessments were made over 2 days to measure the CAR and DCS, and fatigue was measured by the Würzburg Fatigue Inventory for MS (Flachenecker et al., 2006). The Gold et al. (2011) study methodology was described in detail in section 4.1.3.2 of Chapter 4 (p. 132). Gold et al. (2011) found no association between any cortisol outcome and fatigue, but there were several methodological limitations. It was not clear if cortisol assessments were made on weekdays, weekends, or both; an important omission given greater CARs observed on weekdays in previous studies (Kunz-Ebrecht et al., 2004; Schlotz et al., 2004). No objective or self-report measurement of “actual” saliva sampling time was

recorded, so compliance checks were absent and there was no indication of the number of missing samples (Dockray et al., 2008; Kudielka et al., 2003). Potentially-confounding behaviours, such as eating, smoking, physical exercise, or brushing teeth (introducing blood into saliva samples) were not restricted or monitored. Analysis was also not adjusted for the use of DMTs. Fatigue is a known symptom of MDD (American Psychiatric Association, 1994) and some fatigue severity would likely stem from MDD rather than MS; a sensitivity analysis without individuals with MDD would have been informative.

6.1.1 Pharmacological challenge tests and MS fatigue

A study using the standard Dex/CRH test (see Appendix B for standard protocols) has suggested an ACTH hyper-response to CRH stimulation in a fatigued (FSS ≥4) RRMS group compared to a non-fatigued RRMS group (Gottschalk et al., 2005). Two other studies found no association between Dex/CRH test outcomes and fatigue in MS, but recruited all MS-types (RRMS, SPMS, and PPMS) to their sample (Heesen, Gold, et al., 2002; Heesen et al., 2006). Previously, studies have demonstrated greater HPA axis dysfunction (non-suppression by dexamethasone) in PPMS types (53-69%) than RRMS (13-57%) (Heesen et al., 2006; Ysrraelit et al., 2008), meaning heterogeneous MS groups may have masked any association between Dex/CRH outcome and fatigue.

In summary, few studies have examined associations between MS fatigue and HPA axis activity or function, and little clarity is presented by the combined literature. MS-type may be an important moderator of HPA axis activity, and future studies examining CSA in MS should seek to include homogeneous MS samples. Unstimulated CSA in everyday life was not associated with fatigue in RRMS (Gold et al., 2011), but the study had several methodological flaws which have been discussed. A study has demonstrated HPA axis hyper-response to the Dex/CRH test in fatigued pwRRMS compared to non-fatigued pwRRMS (Gottschalk et al., 2005) but it is not clear how this might be reflected in CSA in everyday life.

6.1.2 The CAR and fatigue experience in the upcoming day

Schlotz et al. (2004) and Kunz-Ebrecht et al. (2004) both demonstrated that CARs are greater on working days than weekends, and another study has

demonstrated better coping with daily stress (attenuated emotional responses) on days with larger CARs (Powell & Schlotz, 2012). The CAR may therefore be implicated in preparing the individual for the demands of the upcoming day (Fries et al., 2009; Schulz et al., 1998); however, whether the CAR influences susceptibility to fatigue during the upcoming day has attracted little study.

One study demonstrated a relationship between low waking cortisol and greater fatigue later in the day in a group of 156 adults (age 50–68 years) of mixed health statuses (Adam et al., 2006). Health conditions were diagnosed in many participants (e.g., 41% high blood pressure, 18% diabetes, 12% "psychiatric problems"). MS prevalence was not reported. Salivary cortisol was measured at awakening (S1 cortisol), 30 minutes later and immediately before going to bed over 3 consecutive days. Recalled "physical symptoms/tiredness" were assessed at the end of each day using a set of three items (α = .73) including "Overall today, to what extent did you feel fatigued?" with responses by 5-point VAS ("Not at all" to "Very much"). Results showed higher daily "physical symptoms/tiredness" scores were predicted by lower S1 cortisol level (γ = -0.082, *SE* = 0.030, p = .007), but not the CAR (slope) itself (γ = -0.028, *SE* = 0.033, p = .401). Prior-day "physical symptoms/tiredness" did not predict next-day S1 cortisol (p = .570) or the CAR (p = .115).

Adam et al. (2006) remains the only study to have examined associations between the CAR and upcoming-day fatigue - albeit, measuring combined physical symptoms and tiredness. Sampling compliance was assessed by Medication Event Monitoring System (MEMS) in 52% of participants, and time of awakening was by self-report. Samples were considered non-compliant if MEMS indicated sampling >10 minutes after self-reported awakening/ sampling times. Non-compliance rates were not reported, but non-compliance indicators were included in statistical models. Adam et al. (2006) suggested low S1 cortisol may reduce constraint over pro-inflammatory factors, manifesting in fatigue later in the day. If true, S1 cortisol may be particularly relevant to fatigue in a disease characterised by heightened inflammation such as MS. This hypothesis is yet to be tested.

6.1.3 Change in MS fatigue over time

Several studies have examined change in fatigue severity over prolonged periods of time in MS. Longitudinal studies over 1-3 years in large MS samples

(*n* > 200) have described persistently severe fatigue in 27-47% of pwMS, while 18-25% report no fatigue at any time point (Johansson, Ytterberg, Hillert, Holmqvist, & von Koch, 2008; Lerdal et al., 2007; Téllez et al., 2006). These rates indicate a large number of pwMS experience clinically-significant changes in fatigue severity over periods of up to 3 years; however, understanding of contributing factors is limited. Moderate disease severity (Expanded Disability Status Scale, EDSS, range 4.0–5.5) was associated with change in fatigue severity only when combined with progressive MS-type or >10 years since diagnosis (Johansson et al., 2008), and EDSS was not associated with change in fatigue elsewhere (Téllez et al., 2006).

HPA axis activity may be relevant to change in MS fatigue over the long-term. The only study in the systematic review presented in Chapter 5 to include a follow-up fatigue assessment following salivary cortisol assessments at baseline was in the Whitehall II (civil service) cohort, demonstrating low waking cortisol and flatter DCS predicted concurrent and future (2–3 years later) fatigue (Kumari et al., 2009). The cohort had mixed health-statuses, and CNS medications, cardiovascular disease medications, and depressive symptoms were included as covariates in analysis. The findings suggested attenuated HPA axis activity may predate fatigue; however, it is not clear whether this finding would generalise to MS. A longitudinal examination of the relationship between HPA axis activity and change in MS fatigue severity has yet to be reported.

6.1.4 Chronic stress

MS fatigue is sometimes described as either primary or secondary fatigue, where primary fatigue results from centrally-mediated disease processes including endocrine abnormalities, and secondary fatigue results from other variables including psychological factors (Kos et al., 2008). A review of MS fatigue stated "fatigue in MS does not appear to have a psychological basis [...but...] the experience of fatigue can be influenced by psychological factors" (Kos et al., 2008, p. 95). This latter point was demonstrated in Chapter 3.

There is evidence chronic stress itself can manifest in alterations of HPA axis activity. In a meta-analysis of studies examining chronic (psychological) stress and HPA axis activity, chronic stress was associated with decreased morning cortisol levels, d = -0.08, 95% CI [-0.14, -0.03]; elevated evening samples, d = 0.18, 95% CI [0.09, 0.26]; flatter DCS, d = 0.39, 95% CI [0.18,

0.60]; greater circadian cortisol output, *d* = 0.31, 95% CI [0.20, 0.41]; and less cortisol suppression post-dexamethasone, *d* = -0.23, 95% CI [-0.40, -0.07] (G. E. Miller et al., 2007). If chronic stress moderates any relationship between CSA and MS fatigue, it would evidence psychosocial experience directly impacting what would otherwise be considered a primary source of fatigue stemming from disease processes. It is already well-established that stress can affect MS disease course: stressful life events were associated with risk of relapse in RRMS in a meta-analysis (d = 0.53; Mohr et al., 2004) and have been shown to precede Gd+ lesions on MRI (OR = 1.64; Mohr et al., 2000). If an association between the HPA axis and MS fatigue is moderated by chronic stress, it would contribute towards justifying stress-management interventions for fatigue in RRMS.

6.1.5 Aims of the present study

This chapter tested several hypotheses. Firstly, at baseline (1) a smaller CAR is associated with RRMS fatigue; (2) a flatter DCS is associated with RRMS fatigue; and (3) S1 cortisol is negatively associated with upcoming-day fatigue in RRMS within-subjects. Similar associations between CSA and fatigue were expected in both groups. The final hypothesis was (4) attenuated cortisol variability (CAR AUCi, DCS) is predictive of change in fatigue severity at 6-month follow-up in both groups. If there was evidence for any hypothesis, chronic stress was tested as a potential moderator in an exploratory analysis.

6.2 Method

General methodology was presented in detail in section 2.2.2 of Chapter 2. Relevant features of the EMA protocol and 6-month follow-up are briefly presented below for convenience, followed by a detailed overview of the statistical procedures carried out.

6.2.1 Participants

An RRMS group (*n* = 38) and healthy control group (*n* = 38) were recruited. All individuals in the RRMS group had a neurologist-confirmed RRMS diagnosis, and were in remission-phase (not within 3 months of a relapse). Participant characteristics were presented in Table 2 in Chapter 2 (p.51).

6.2.2 Design

The procedure undertaken by each participant was presented in detail in section 2.2.2 of Chapter 2. After the introductory session, 4 consecutive weekdays of EMA was completed. Two facets of the EMA design were relevant to this chapter: (1) the fixed-occasion event-based morning design (events S1–S3) measuring the CAR, and (2) the variable-occasion quasi-random design (events A1–A6, 1000h–2000h) measuring the DCS and other momentary assessments. The end-of-day event (2100h) was not relevant to this chapter. After 6 months, follow-up questionnaires were posted to participants with return by pre-paid mail.

6.2.3 Measures

6.2.3.1 Baseline measures

Baseline measures relevant to this chapter were: (1) the self-administered Expanded Disability Status Scale (EDSS; Bowen et al., 2001) measuring neurological symptoms in the RRMS group; (2) the Fatigue Scale (FS; Chalder et al., 1993) measuring fatigue severity; (3) the Modified Fatigue Impact Scale (MFIS-5; Ritvo et al., 1997) measuring the impact of fatigue; (4) the Chronic Stress Screening Scale (CSSS; Schulz et al., 2004) measuring chronic exposure to stress; and (5) the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D; Zigmond & Snaith, 1983) measuring depressive symptoms.

The FS was used as both a continuous scale and to categorise individuals in the RRMS group as with or without "clinically meaningful fatigue" based on the ≥4 cut-off on the bimodal scoring system (Chalder et al., 1993). Demographic and other background variables were assessed as potential covariates. Further details of baseline measures, their validity and justifications were presented in section 2.2.3.1 of Chapter 2.

6.2.3.2 Ecological momentary assessments

6.2.3.2.1 Cortisol

Salivary cortisol assessments were provided with each auditory prompt upon awakening and 30min and 45min after awakening (S1-S3 events) to measure the CAR, and the variable-occasion design between 1000h and 2000h

(A1-A6 events) to compute the DCS. Three facets of the CAR were measured: (1) AUCg, (2) AUCi, and (3) S1 cortisol (Clow et al., 2010; J. C. Pruessner et al., 2003). Formulae for the CAR AUCg and AUCi markers are presented in Appendix G. The AUCg represents the estimated total cortisol output, the AUCi represents the change in cortisol output post-awakening, and S1 cortisol marks the end of the pre-awakening rise in cortisol (Clow et al., 2004). The DCS was computed as the linear slope parameter from the 3-level multilevel model of natural log-transformed cortisol obtained from A1 to A6 events.

6.2.3.2.2 Self-report

Sleep quality and length of sleep was measured by self-report at S1. At each quasi-random event (A1–A6), behavioural activities (physical activity, drinking coffee, eating a meal, and smoking) were assessed by check-box response (Yes/No). Momentary fatigue was also measured at events A1-A6 using the single-item "How much fatigue (tiredness, weariness, problems thinking clearly) do you feel right now?" Information about the development of this item was presented in section 2.2.3.2.3.3 of Chapter 2. Momentary stressors and mood measured at events A1-A6 were not relevant to this chapter.

6.2.4 Follow-up measures

At follow-up, participants completed the following questionnaires: FS, MFIS-5, CSSS, and EDSS. Participants also indicated change in MS diagnosis, comorbidity diagnosis, medications, number of MS relapses since baseline, and change in employment status.

6.2.5 Statistical analysis

MLMs were used to test hypotheses 1-3. Hypothesis 4 was tested with stepwise linear regression models with follow-up FS score as outcome. Specified models use notation from Snijders and Bosker (2012). All statistical analysis was conducted using SPSS Version 20.0 and some figures were produced using SigmaPlot Version 12.3. An alpha level of .05 was used for all statistical tests.

6.2.5.1 CAR and baseline fatigue

To test hypothesis 1, a set of 2-level multilevel models of CAR outcomes were built as presented previously in Equation 5 in section 4.2.5.2 (p.143). S1 cortisol was natural-log transformed to correct for positively-skewed distributions (Schlotz, 2011); transformation was not necessary for AUC outcomes. As was the case in Chapter 4, no covariates were retained in the final model of the CAR. A compound symmetry variance covariance matrix was used as best-fit for S1 cortisol and AUCg outcomes, and an autoregressive matrix for AUCi (Appendix P, Tables A-C). To test the relationship of CAR with baseline fatigue across groups, FS/MFIS score and group (RRMS, Control) were entered as fixed effects at Level-2, with a group by FS/MFIS interaction indicating group differences in the association. A CAR "responder" analysis was performed based on the ≥1.5nmol/L response criterion (R. Miller et al., 2013).

6.2.5.2 Diurnal cortisol slope and baseline fatigue

Hypothesis 2 was tested with a 3-level multilevel model (Equation 6 in section 4.2.5.3, p.145) nesting 6 assessments (A1-A6; Level-1) within 4 days (Level-2) within individuals (Level-3). Cortisol was natural-log transformed to correct for positively-skewed distributions (Schlotz, 2011). The model contained the fixed effect of time (centred at 1000h) at Level-1, and time effects were permitted to vary (random effects) at Level-3. As was the case in chapter 4, recent meal eating and smoking (Level-1 variables) were the only covariates retained in the final model, and were added as fixed effects to the specified growth model in Equation 6. An unstructured variance-covariance matrix was used to get an estimation of the covariance of intercept and time random effects. To test the relationship with fatigue across groups, FS/MFIS score and group were entered as fixed effects at Level-3, with a group by FS/MFIS interaction effect testing group differences in the association.

6.2.5.3 Clinically meaningful fatigue

In secondary analyses, the RRMS group was split into those with (RRMS-f, n = 21) and without (RRMS-nf, n = 17) "clinically meaningful fatigue" by the FS ≥4 cut-off. For group comparisons, specified multilevel models for CAR outcomes and DCS were rerun using dummy variables to represent the three groups as fixed effects (RRMS-f, RRMS-nf, control).

6.2.5.4 Daily CAR and fatigue associations

A multilevel model with daily fatigue severity as outcome was built to test hypothesis 3. Daily fatigue severity was operationalized by the daily momentary fatigue mean. The MLM nested four daily fatigue ratings (Level-1) within 76 individuals (Level-2), and the null model is presented in Equation 8.

Equation 8

Multilevel model for the daily fatigue severity

$$Daily\ Fatigue\ Severity_{di} = \gamma_{00} + U_{0i} + R_{di}$$

Where $R_{di} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$

Where $Daily\ Fatigue\ Severity_{di}$ is the value of the outcome for person i on day d. The average intercept is indicated by γ_{00} and indicates the average outcome across all participants. Random effects at Level–2 (individual) are indicated by U with U_{0i} denoting the deviation of the intercept for person i from the average intercept. R_{di} indicates the residuals for each unique assessment.

Different variance covariance matrices were examined for best fit (Singer & Willett, 2003) and the compound symmetry matrix was most efficacious (see Appendix W). In respective models, each CAR predictor (S1 cortisol, AUCg, AUCi) was person-centred and entered as a fixed effect at Level-1 along with group and effect by group interaction effects at Level-2.

6.2.5.5 Chronic stress as a moderator

Where a statistically significant association between cortisol and fatigue was found, further models were tested including CSSS score as a fixed effect at the individual level to examine whether chronic stress moderated the association. Subsequently, HADS-D score, employment status, age, and gender were all entered in sensitivity analyses to check for extraneous confounders. For the RRMS group only, sensitivity analyses were performed with EDSS score and DMT-use as fixed effects.

6.2.5.6 Longitudinal examination of cortisol and fatigue

Hypothesis 4 was tested by stepwise multiple regression, with follow-up FS score as outcome. Predictors were entered in the following steps: (1) baseline FS score and group; and (2) person-mean CAR marker (S1, AUCg, or AUCi) with a computed CAR by group interaction term. Analysis using change scores for fatigue was not performed as baseline values can be negatively correlated with change; those with low scores generally improve more than those with high scores (regression to the mean) (Vickers & Altman, 2001).

A recent study found CAR predicted future progression in neurological disability measured by the EDSS (Kern et al., 2013), so a supplementary analysis was performed seeking to replicate this finding. Follow-up EDSS score was outcome, with stepwise entering of predictors into multiple regression: (1) baseline EDSS score and group; and (2) person-mean CAR marker (S1, AUCg, or AUCi) with CAR by group interaction.

6.3 Results

At baseline, FS scores were significantly higher in the RRMS group (M_f = 17.58 ±7.09, range 1 – 32) than the healthy control group (M_f = 11.55 ±2.87, range 4 – 21, $p < .001$). MFIS-5 scores were also higher in the RRMS group (M_f = 9.39 ±4.25, range 0 – 16) than the healthy control group (M_f = 3.87 ±3.82, range 0 – 14). Group differences represented large effect sizes for both the FS, d = 1.11, 95% CI [0.62, 1.58], and MFIS-5, d = 1.36, 95% CI [0.85, 1.85].

6.3.1 Outliers and assumptions

Appendix N presents scatterplots of CAR measures (AUCg, AUCi, S1, S2, and S3) and transformed daytime cortisol assessments (A1-A6) against participant ID for outlier identification. Sensitivity analyses (without outliers) were performed for each hypothesis test where appropriate and reported below. Residual distributions from models used to test hypotheses did not substantially deviate from normality (see Appendix V for residual histograms).

6.3.2 Cortisol awakening response and baseline fatigue

Compliant CARs were provided on 258 of 304 (84.8%) assessment days. One RRMS participant was excluded from the CAR dataset as they provided no compliant CARs. No participants were excluded on missingness criteria for the S1 analysis. CAR AUCg was not associated with FS or MFIS measures in either the RRMS group (all *p*s > .104) or control group (all *p*s > .725). However, there were statistically significant associations of S1 cortisol and CAR AUCi with FS scores; but only in the RRMS group (summarised in Table 23).

Table 23 Multilevel Parameter Estimates of S1 Cortisol and CAR AUCi as a Function of Fatigue measures for RRMS and Control groups

Fixed effects	RRMS	Control	Group difference
S1 Cortisol	γ (*SE*)	γ (*SE*)	γ (*SE*)
FS Total	-0.022 (0.008)**	-0.008 (0.019)	-0.014 (0.021)
FS Physical	-0.028 (0.012)*	-0.009 (0.024)	-0.019 (0.027)
FS Mental	-0.060 (0.021)**	-0.011 (0.059)	-0.049 (0.063)
MFIS-5	-0.024 (0.014)˜	0.002 (0.015)	-0.026 (0.020)
CAR AUCi			
FS Total	0.151 (0.058)*	0.026 (0.146)	0.125 (0.157)
FS Physical	0.216 (0.093)*	0.019 (0.176)	0.206 (0.195)
FS Mental	0.392 (0.175)*	0.122 (0.459)	0.195 (0.486)
MFIS-5	0.090 (0.105)	0.048 (0.113)	0.042 (0.154)

Note. Model intercept and group effects not shown. Data presented is from "responder" and "non-responder" days. FS represents Fatigue Scale; MFIS-5, Modified Fatigue Impact Scale; S1, sample upon awakening; CAR AUCi, cortisol awakening response area under the curve increase. ** $p < .01$, * $p < .05$, ˜ $p < .10$.

S1 cortisol was negatively associated with FS score in the RRMS group, p = .007, 95% CI [-0.038, -0.006], such that lower S1 cortisol levels were associated with greater fatigue (see Figure 17), but this relationship was not present in the control group, p = .700, 95% CI [-0.046, 0.031]. Excluding potential outliers in S1 cortisol (see Appendix N, Figure C) did not change the result in the RRMS group, γ = -0.022, *SE* = 0.008, p = .007, 95% CI [-0.036, 0.006]. Excluding a potentially influential outlier at FS = 1 did not affect the result, γ = -0.020, *SE* = 0.009, p = .026, 95% CI [-0.037, -0.002]. In the RRMS group, S1 cortisol was relevant to both physical fatigue, p = .016, 95% CI [-0.051, -0.005], and mental fatigue, p = .005, 95% CI [-0.102, -0.018], but the association with fatigue impact (MFIS-5) did not quite reach statistical significance, p = .081, 95% CI [-0.051, 0.003].

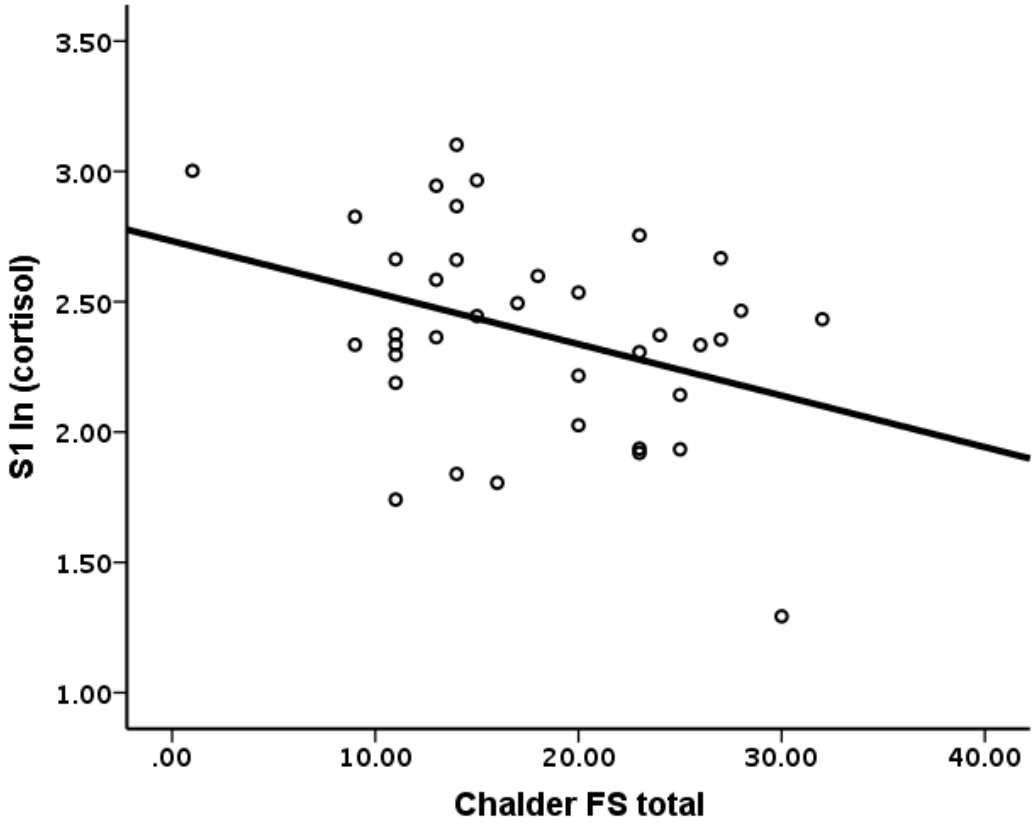


Figure 17. Scatter plot of person-aggregated S1 cortisol levels against baseline FS score in RRMS group only.

Higher fatigue scores in the RRMS group were also associated with larger CAR AUCi, p = .017, 95% CI [0.029, 0.285], (see Figure 18). This result was not substantially influenced by potential CAR AUCi outliers (see Appendix N, Figure B), γ = 0.15, SE = 0.06, p = .011, 95% CI [0.036, 0.267], nor by the potential outlier on the FS, γ = 0.12, SE = 0.06, p = .062, 95% CI [-0.006, 0.246]. CAR AUCi was associated with both physical fatigue, p = .008, 95% CI [0.031, 0.400], and mental fatigue, p = .028, 95% CI [0.043, 0.739]. There was no association between fatigue impact (MFIS-5) and CAR AUCi in the RRMS group, p = .389, 95% CI [-0.128, 0.326], and no association between the CAR AUCi and any fatigue measure in the control group (all ps > .676).

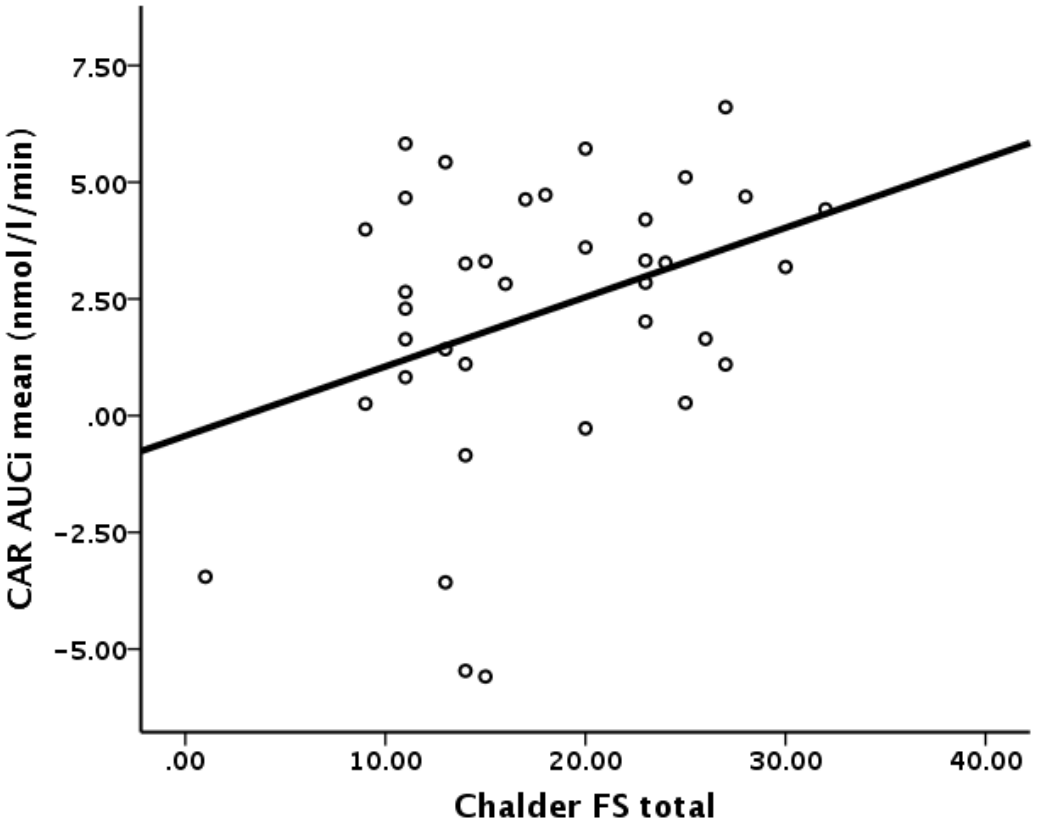


Figure 18. Scatter plot of person-aggregated cortisol awakening response AUCi against baseline FS score in RRMS group only.

After the addition of CSSS score as a fixed effect with group*CSSS interaction to the model, the association of FS score with S1 cortisol, $\gamma = -0.022$, $SE = 0.008$, $p = .006$, 95% CI [-0.038, -0.006], and FS score with CAR AUCi, $\gamma = 0.16$, $SE = 0.06$, $p = .014$, 95% CI [0.03, 0.29], remained, indicating that chronic stress did not moderate the association between fatigue and cortisol in RRMS. When other potentially explanatory variables were added as fixed effects (EDSS score, DMT use, HADS-D score, age, and gender) there were still no substantial changes in the relationship of FS score with S1 cortisol ($p = .020$) nor CAR AUCi ($p = .065$).

6.3.2.1 CAR and clinically meaningful fatigue

Table 9 in Chapter 3 (p.108) summarised RRMS-f, RRMS-nf, and control groups for key variables. There was no difference between groups for age, gender, employment, average sleep quality, or average sleep duration. Although the RRMS-f subgroup were somewhat more disabled (higher EDSS) than the RRMS-nf group, this was not to statistical significance.

Figure 19 depicts the subgroup mean CARs. S1 cortisol was elevated in the RRMS-nf subgroup compared to both the RRMS-f subgroup, $\gamma = 0.23$, $SE = 0.11$, $p = .049$, 95% CI [0.001, 0.46], and the healthy control group, $\gamma = 0.27$, $SE = 0.10$, $p = .009$, 95% CI [0.07, 0.48]. There was no difference for S1 cortisol between RRMS-f and healthy control groups, $\gamma = 0.05$, $SE = 0.10$, $p = .626$, 95% CI [-0.14, 0.24]. Using person-aggregated data based on t-tests, the difference in S1 cortisol between RRMS-f and RRMS-nf represented a large effect size, $d = -0.96$, 95% CI [-1.61, -0.26].

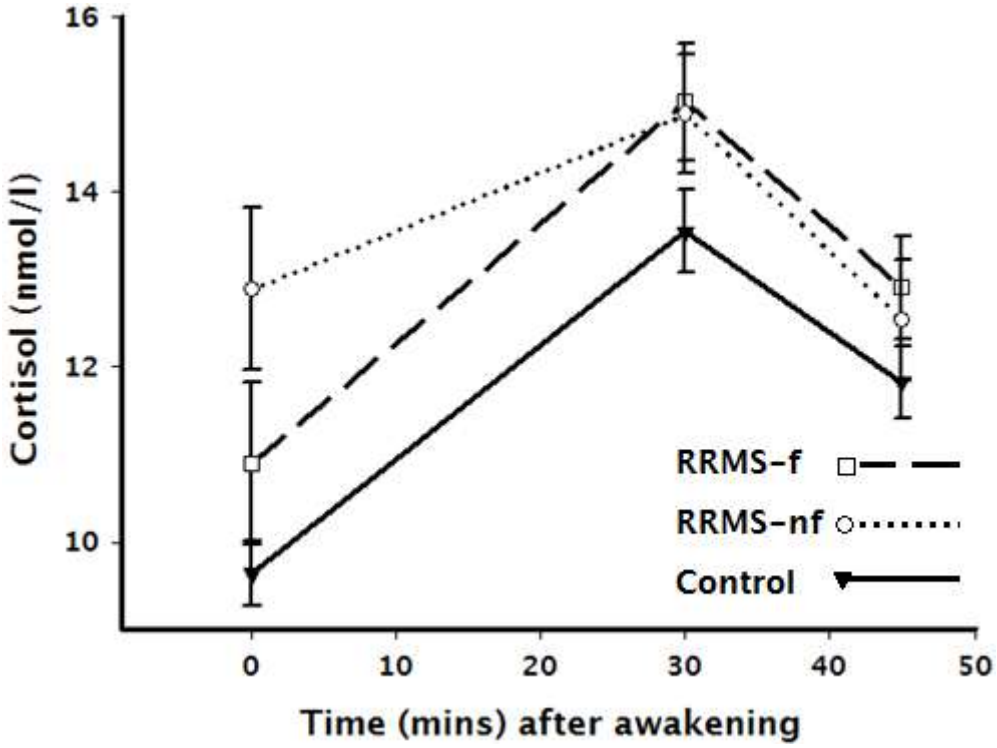


Figure 19. Cortisol awakening response represented by the mean of the within-subject means for samples at 0 (S1), 30 (S2), and 45 minutes (S3) post-awakening. Error bars represent the standard error of the mean.

There was also a difference for CAR AUCi between RRMS-f and RRMS-nf subgroups, although not quite to statistical significance, $\gamma = 1.69$, $SE = 0.85$, $p = .052$, 95% CI [-0.01, 3.38], such that CAR AUCi was, on average, 1.69 nmol/L/min greater in the RRMS-f group than the RRMS-nf subgroup and represented a medium effect size with aggregated data, $d = 0.51$, 95% CI [-0.15, 1.15]. The average CAR AUCi for the control group was no different to the RRMS-f, $\gamma = -0.57$, $SE = 0.70$, $p = .416$, 95% CI [-0.82, 1.96], or RRMS-nf subgroups, $\gamma = 1.11$, $SE = 0.78$, $p = .156$, 95% CI [-2.66, 0.43].

CAR AUCg was greater in the RRMS-nf subgroup than control group, $\gamma = 2.17$, $SE = 0.95$, $p = .025$, 95% CI [0.28, 4.07], such that individuals in the RRMS-nf group produced, on average, 2.17 nmol/L/min more cortisol over the 45-minute post-awakening period than those in the control group. Differences between RRMS-f and RRMS-nf subgroups, $\gamma = 1.14$, $SE = 1.05$, $p = .283$, 95% CI [-.3.24, 0.96], and RRMS-f and healthy control groups, $\gamma = 1.03$, $SE = 0.86$, $p = .237$, 95% CI [-0.69, 2.75], were not statistically significant. Subgroup differences for all three CAR measures are presented in Figure 20.

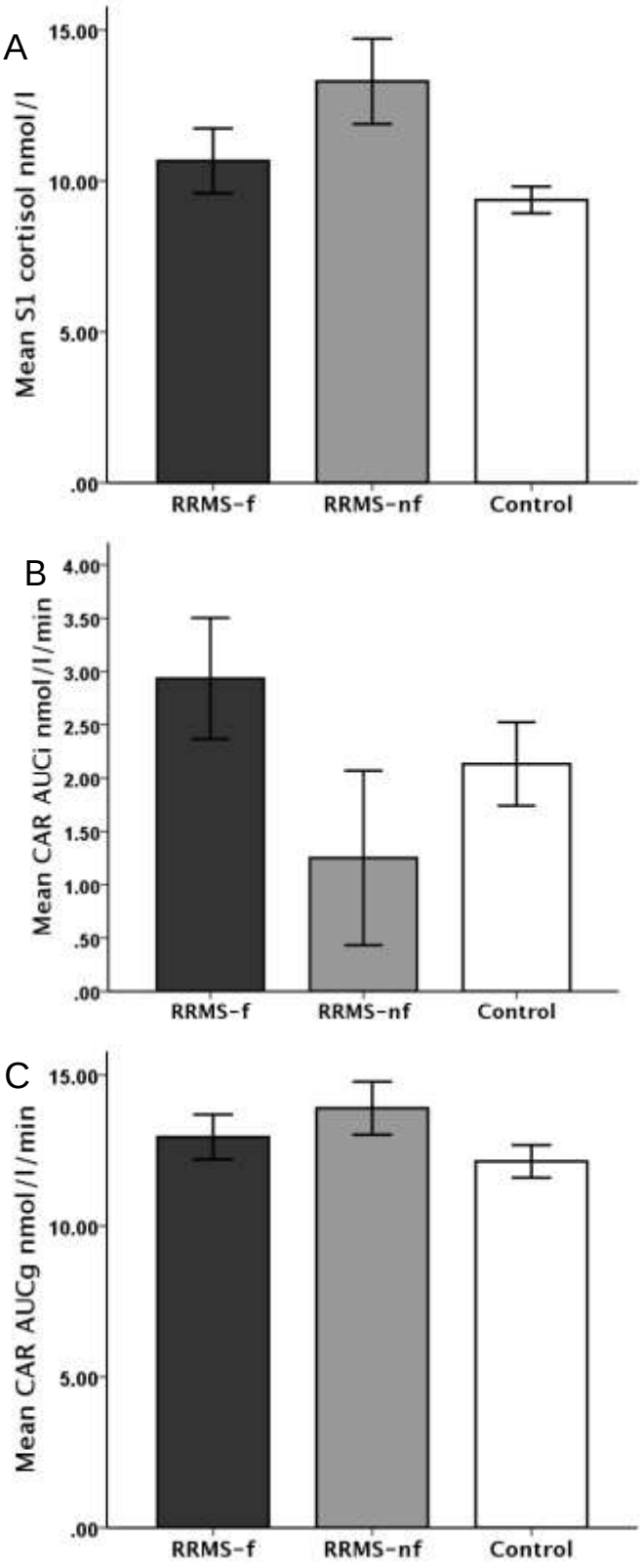


Figure 20. Bar charts of cortisol awakening response markers for each group: (A) untransformed S1 cortisol; (B) AUCi; (C) AUCg. Error bars indicate standard error of the mean.

6.3.2.2 **"Responder" analysis**

There were 177 out of 258 (68.6%) CAR measurement days classified as "responder" days on the 1.5nmol/L criterion (R. Miller et al., 2013), with similar prevalence in the RRMS group (69.7%) and the control group (67.5%). However, non-response rates were far higher in the RRMS-nf group (42.8%) than both the RRMS-f group (78.5%) and control group (67.5%). Six individuals did not provide a "responder" CAR on any of the 4 assessment days.

In "responder"-only analyses, the association of FS with S1 cortisol was diminished such that the effect was no longer significant, $\gamma = 0.007$, $SE = 0.008$, $p = .382$, 95% CI [-0.024, 0.009], similar to the association with CAR AUCi, $\gamma = -0.009$, $SE = 0.072$, $p = .902$, 95% CI [-0.153, 0.135].

6.3.3 Diurnal cortisol slope and baseline fatigue

Non-response, missing samples, or incorrectly labelled random codes left 1637 out of 1824 (89.7%) DCS samples for analysis. Eight assessment days belonging to eight individuals (five RRMS; three control) contained less than three cortisol assessments, and so were excluded from the analysis. All participants provided more than 2 valid assessment days.

The DCS was not associated with FS score in the RRMS group, $\gamma = 0.001$, $SE = 0.001$, $p = .291$, 95% CI [-0.0009, 0.0030]. However, flatter DCS was associated with higher FS scores in the control group, $\gamma = 0.004$, $SE = 0.002$, $p = .030$, 95% CI [0.0004, 0.0085]. DCS was associated with physical fatigue in the control group, $\gamma = 0.005$, $SE = 0.003$, $p = .038$, 95% CI [0.0003, 0.0103], but the association with mental fatigue did not reach statistical significance, $\gamma = 0.010$, $SE = 0.006$, $p = .117$, 95% CI [-0.0025, 0.0221].

DCS comparisons of RRMS-f, RRMS-nf, and control subgroups are presented in Figure 21. Cortisol levels were elevated in the RRMS-nf group at 1000h compared to the control group, $\gamma = 0.19$, $SE = 0.09$, $p = .044$, 95% CI [0.005, 0.378], but the comparison with the RRMS-f subgroup did not reach statistical significance, $\gamma = 0.15$, $SE = 0.11$, $p = .160$, 95% CI [-0.060, 0.358]. Therefore, despite cortisol levels being similar at S3 for all groups (see Figure 19), cortisol level decreased more slowly from the S3 event to 1000h in the RRMS-nf subgroup. This effect remained after controlling for awakening time.

There were no differences for DCS between subgroups (all *p*s > .319), and no differences in 2000h cortisol (all *p*s > .548).

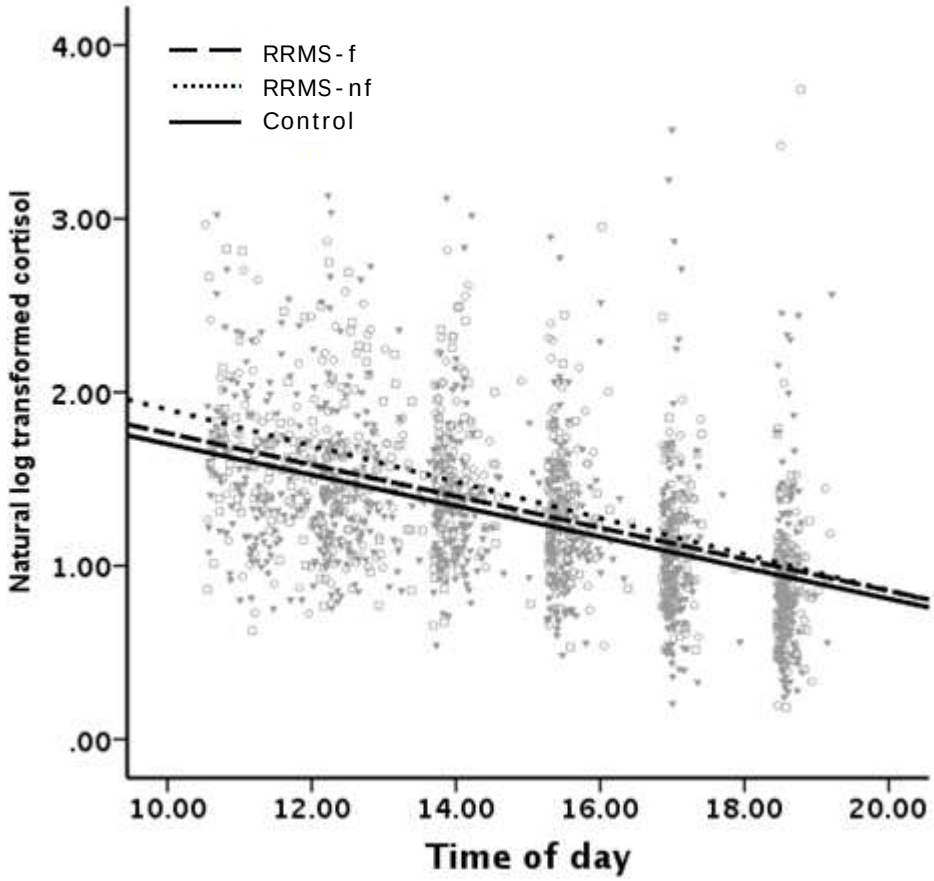


Figure 21. Average diurnal cortisol slopes comparing relapsing-remitting multiple sclerosis groups (fatigued and not fatigued) and healthy control group.

Note. Small triangles represent healthy control assessments; squares represent RRMS-f; circles represent RRMS-nf.

6.3.4 CAR and fatigue in the upcoming day: within-subjects analysis

No associations between person-centred CAR measures and upcoming-day fatigue were statistically significant. S1 cortisol, p = .63, 95% CI [-0.78, 0.47]; CAR AUCi, p = .91, 95% CI [-0.07, 0.07]; and CAR AUCg, p = .81, 95% CI [-0.06, 0.08) were not associated with upcoming-day fatigue in the RRMS group. Although not quite to statistical significance, greater CAR AUCg was

associated with greater levels of latter day-fatigue in controls, within-subjects, $\gamma = 0.08$, $SE = 0.04$, $p = .054$, 95% CI [-0.001, 0.163]. S1 cortisol, $p = .50$, 95% CI [-0.46, 0.95], and CAR AUCi, $p = .90$, 95% CI [-0.08, 0.07], were not associated with upcoming-day fatigue in the control group.

6.3.5 CAR and follow-up fatigue and neurological symptoms

Follow-up reports were provided by 79% (30/38) of the RRMS group and 76% (29/38) of the control group. One individual in the RRMS responded 0 to all questions on all follow-up measures and was excluded from the follow-up analysis. One individual in the RRMS group was diagnosed with osteoarthritis before follow-up, four individuals (two RRMS, two control) had become pregnant, two individuals (one RRMS, one control) were now taking iron supplements, and two individuals in the control group were diagnosed with MDD (one also with anxiety). These individuals no longer met eligibility criteria and were excluded from the follow-up analysis, leaving 25 individuals in the RRMS group and 24 in the control group. Five individuals in the RRMS group (20%) reported a change in DMT regimen, including two who reported stopping all medications. These individuals were not excluded, but sensitivity analysis was conducted for any statistically significant effects. The 25 pwRRMS included at follow-up did not differ from those who were lost on age, duration of illness, fatigue, EDSS, or any cortisol measure (*p*s >.05).

Of those in the RRMS group fatigued at baseline (scored ≥4 on FS bimodal score) and who subsequently completed follow-up, 82% (14/17) remained fatigued after six months. Of those in the RRMS group not fatigued at baseline, only one individual (1/8) reported fatigue at follow-up. Baseline fatigue was a highly significant predictor of follow-up fatigue, $B = -0.66$, $SE = 0.09$, $t(46) = 7.38$, $p < .001$, 95% CI [0.48, 0.85], accounting for 53% of variance. Controlling for baseline fatigue severity, in the RRMS group, there was no effect of average S1 cortisol, $B = -1.00$, $SE = 1.71$, $t(45) = -0.56$, $p = .56$, 95% CI [-4.45, 2.45]; CAR AUCg, $B = -0.14$, $SE = 0.20$, $t(44) = -0.68$, $p = .50$, 95% CI [-0.53, 0.26]; or CAR AUCi, $B = 0.01$, $SE = 0.27$, $t(44) = 0.04$, $p = .97$, 95% CI [-0.53, 0.55], on FS score at follow-up. There were no statistically significant associations in the control group, nor any group*effect interactions (all *p*s > .15).

EDSS score at baseline was a highly significant predictor of follow-up EDSS score, B = 0.85, SE = 0.10, $t(23)$ = 8.16, p < .001, 95% CI [0.63, 1.06], accounting for 73% of variance. Controlling for baseline EDSS, there was no effect of baseline average S1 cortisol, B = -0.13, SE = 0.30, $t(22)$ = -0.42, p = .68, 95% CI [-0.75, 0.50]; CAR AUCg, B = -0.03, SE = 0.04, $t(22)$ = -0.92, p = .37, 95% CI [-0.11, 0.04]; or CAR AUCi, B = 0.01, SE = 0.04, $t(22)$ = 0.37, p = .72, 95% CI [-0.07, 0.09], on EDSS score at follow-up.

6.3.6 Results summary

In the RRMS group, high levels of fatigue severity were associated with low S1 cortisol level as well as greater cortisol responses to awakening (CAR AUCi). The DCS was not associated with fatigue in the RRMS group. Although the CAR was not associated with fatigue in the control group, a flatter DCS was associated with fatigue. In a within-subjects analysis, greater CAR output (AUCg) was associated with greater upcoming-day fatigue in controls, although not quite to statistical significance. No within-subject associations between CAR outcomes and fatigue were found in the RRMS group. There was no association between baseline cortisol measures and follow-up fatigue in either group.

6.4 Discussion

The chapter explored the relationship between everyday life CSA and fatigue experience in an RRMS group without depressive comorbidity. Results demonstrated a robust relationship between the CAR and fatigue in RRMS such that low S1 cortisol (end of the pre-awakening rise) and accentuated AUCi (post-awakening rise) were associated with greater baseline fatigue. Psychological factors (chronic stress, depression) did not moderate the relationship between the CAR and fatigue, suggesting the effect is more likely a manifestation of underlying disease processes than of psychological cause. Individuals in the control group with greater fatigue at baseline had flatter diurnal slopes, supporting findings in other general population studies (Kumari et al., 2009; Lindeberg et al., 2008), but this association was not present in the RRMS group.

The CAR was not associated with latter-day fatigue in RRMS, but a within-subjects positive association in the control group between AUCg and later

fatigue was close to statistical significance. This potential relationship requires replication across more sampling days, but implies flexibility in the CAR in healthy individuals which may elicit effects over fatigue experience throughout the rest of the day. CSA did not predict change in fatigue severity 6 months later in either group.

6.4.1 Cortisol and fatigue in RRMS

This chapter adds to existing literature examining the relationship between HPA axis activity and fatigue in RRMS (Gold et al., 2011; Gottschalk et al., 2005; Heesen, Gold, et al., 2002; Heesen et al., 2006). The findings support previous evidence of HPA axis hyperactivity in RRMS in association with fatigue (Gottschalk et al., 2005). While Heesen, Gold, et al. (2002) and Heesen et al. (2006) did not observe an association between Dex/CRH test outcomes and fatigue, their populations contained both RRMS and progressive MS types, suggesting a different physiological basis for fatigue in RRMS compared to progressive MS types. This assertion requires further investigation.

Gold et al. (2011) found the CAR (AUCg) was not associated with RRMS fatigue and, while the present study found no direct relationship between fatigue severity and AUCg, it did demonstrate AUCg and S1 cortisol were elevated only in those pwRRMS who did not report "clinically meaningful fatigue". Considered alongside the increased cortisol output at 1000h in the RRMS-nf group, it would seem plausible that enhanced cortisol output during the morning is an adaptive mechanism in RRMS, possibly in response to the heightened levels of circulating pro-inflammatory cytokines seen in RRMS (Heesen et al., 2005; Ysrraelit et al., 2008). Speculatively, one may suggest that MS fatigue manifests in the absence of sufficient down-regulation of inflammation by cortisol, particularly in the mornings.

The most robust finding was an association of reduced S1 cortisol and elevated CAR AUCi with fatigue in RRMS, including lower S1 cortisol and higher AUCi in those with "clinically meaningful fatigue". The role of elevated CAR AUCi contradicts the conclusion of the systematic review (Chapter 5) of generally attenuated diurnal cortisol variability (measured by markers such as AUCi and DCS) in association with fatigue across populations. This conclusion would seem not to generalise to MS. Of note, the systematic review contained a

study in rheumatoid arthritis (Dekkers et al., 2000) demonstrating a positive association between the cortisol response to awakening (slope from waking to peak) and fatigue severity. The fact rheumatoid arthritis is, like MS, an autoimmune disease characterised by high levels of inflammation may indicate different mechanisms of fatigue in this category of disease compared to chronic fatigue syndrome.

That the CAR did not predict change in fatigue at 6-month follow-up may have been due to a lack of variability from baseline to follow-up, demonstrated by only four individuals making "clinically meaningful" changes in fatigue severity. The test-retest correlation for the FS from baseline to follow up was *rho* = .59. It would be beneficial for future studies to include more than one follow-up point and over a longer period. The observation that fatigue did not change much over 6 months is potentially interesting, and indicates that, although there is a lot of variability in fatigue severity within days (Chapter 3), trait-like fatigue severity appears relatively stable over longer periods.

Although the DCS was not associated with fatigue in RRMS in this study, the fact flatter DCS was associated with fatigue in the control group corroborated well with the findings of the systematic review (Chapter 5). The cause of flat DCS is unknown, although it has previously been associated with chronic stress (G. E. Miller et al., 2007). In the present study, chronic stress did not moderate the relationship in controls. Taken together, the findings for the RRMS and control groups would suggest different mechanisms underlying associations between CSA and fatigue, with healthy individuals demonstrating a relationship similar to that seen in chronic fatigue syndrome whereas, in MS, inflammatory pathology may underlie the relationship.

Low S1 cortisol may produce a state whereby inflammation is not adequately controlled in MS, leading to adaptive hyper-responses to awakening. There is evidence of associations between pro-inflammatory cytokines (such as TNF-α and IFN-γ) and fatigue in MS. A study has demonstrated greater TNF-α in pwMS (19 RRMS, 8 SPMS) with fatigue (FSS ≥4, *n* = 26) compared to those without (*n* = 11) (Flachenecker et al., 2004). This has been corroborated by Heesen et al. (2006) in another heterogeneous group of pwMS (17 RRMS, 10 SPMS, 3 PPMS) with greater TNF-α in those with fatigue (FSS ≥5;*n* = 15) than those without (*n* = 15), *d* = 1.20, 95% CI [0.39, 1.94]. Heesen et al. (2006) also demonstrated elevated IFN-γ in the fatigued group,

d = 0.74, 95% CI [0.00, 1.48]; however elevated IFN-γ was not associated with fatigue by Flachenecker et al., (2004) possibly due to the more liberal fatigue criterion used. A recent longitudinal study of 80 rheumatoid arthritis patients over 6 months also implicated pro-inflammatory cytokines in fatigue, as levels of IFN-γ and IL-1β predicted fatigue 1 month later (Evers et al., 2013). Cortisol was measured, but only once between 0900h and 1000h each month, with no reference to time of awakening; lower cortisol was associated with greater fatigue 1 month later, but not quite to statistical significance (γ = -0.075, *SE* = 0.043, *p* = .08). The role of inflammation in MS fatigue and the related implications for CSA requires further research.

There have been suggestions that the inhibitory actions of the hippocampus are most prominent prior to awakening during REM sleep (Clow et al., 2010). Damage or deficiencies in the hippocampus, which have been observed in previous studies in MS (Sicotte et al., 2008) may drive elevated S1 cortisol typically seen in MS and, speculatively, the absence of such a relationship may be associated with fatigue. There is currently no research looking at the relationship between hippocampal damage and fatigue in MS, but this could be a target for future research along with the role of S1 cortisol and CAR AUCi in any relationship.

The results of the within-subjects analysis suggested day-to-day variability in the CAR may affect latter-day fatigue experience in healthy individuals but not pwRRMS. The fact greater cortisol output (AUCg) was associated with greater fatigue was surprising given the systematic review had described output as having limited relevance to fatigue across populations. However, previous studies have been limited to performing between-subjects analyses, and it is possible that within-subjects associations are different. J. Hellhammer et al. (2007) described marked intra-individual state variation in the CAR, yet state variation is relatively infrequently examined (Law, Hucklebridge, Thorn, Evans, & Clow, 2013). Adam et al. (2006) found low S1 cortisol associated with greater "physical symptoms/tiredness" later in the same day. This result was not supported by the data presented in this chapter, although the measure in the Adam et al. (2006) study was not a "true" measure of fatigue and included other symptoms such as headache. Despite apparent contradictions, investigating state variation in the CAR and day-to-day changes in fatigue is a promising direction for future research.

6.4.2 Strengths and limitations

The present study had more saliva assessments over twice as many assessment days than the only other examination of everyday life CSA and MS fatigue (Gold et al., 2011), providing more reliable parameter estimates for the CAR and DCS and greater statistical power to detect effects. The CAR, in particular, has been shown to require several days to provide a reliable estimation given the degree of state variation (J. Hellhammer et al., 2007). Many steps were taken to maximise compliance, including using random code matching of salivette labels and handheld presentations, and electronic prompting of samples. Compliance rates were excellent.

Previous research had presented a confusing overview of the relationship between HPA axis function and fatigue in MS, particularly where populations were heterogeneous. The study addressed this by recruiting only pwRRMS in remission to the clinical group. Therefore, the results here cannot necessarily be generalised to PPMS or SPMS, or individuals experiencing relapse-phase RRMS. Both 24h urinary cortisol and pro-inflammatory cytokines have been shown to be elevated during relapse-phase compared to when in remission (Ysrraelit et al., 2008). However, it would not be possible to examine unstimulated CSA while pwRRMS were treated for relapse with corticosteroids.

The study lacked an objective observation of awakening, so it was not completely certain participants promptly provided S1 samples upon awakening; however, several steps were taken to maximise compliance, including waking by alarm. It could be argued that the difference in S1 cortisol between RRMS-f and RRMS-nf subgroups was merely an artefact of one group being more compliant than the other. There was no reason to believe this was the case, and is considered unlikely; one would probably expect more delays (and artificially inflated S1 cortisol levels) in those who were fatigued than those who were not. In addition, elevated S1 cortisol due to delay would also manifest as an earlier CAR peak, which was not detected.

The study reported a relatively high attrition rate from baseline to 6 months. Although a reasonable proportion (78%) completed and returned the follow-up questionnaires, after re-application of the exclusion criteria, only 65% of the original sample provided data for follow-up. In future studies, it will be important to recruit a larger sample with the expectation that a proportion

will develop comorbidities or other changes will occur, such as pregnancy, resulting in exclusion.

6.4.3 Conclusion

This chapter presented evidence of an association between the CAR and fatigue in RRMS, such that lower awakening (S1) cortisol level and an enhanced awakening response (AUCi) were associated with greater baseline fatigue. There was no evidence of an association between DCS and fatigue in RRMS, but flatter slopes were associated with fatigue in the healthy control group. Although CSA was not associated with change in fatigue 6 months later in either group, this finding should be interpreted with caution due to small numbers in the follow-up analysis and minimal change in fatigue detected. Cortisol appears to have a role in fatigue experienced in RRMS, but whether it is a causal factor remains unclear. Results contrast to blunted CARs typically observed in chronic fatigue syndrome (described in Chapter 5), which may be due to the enhanced neuro-inflammation characteristic of MS.

Chapter 7: General Discussion

The final chapter concludes this thesis by considering its overall contribution to the understanding of fatigue and cortisol secretory activity (CSA) in MS. The first section summarises the work presented in this thesis, identifying the key findings of each chapter and stating the unique contributions to the research literature. These key findings are then critically discussed with relation to the current literature and their implications. Consideration is then given to the generalisability of findings from the clinical study. Finally, an overview of the strengths and limitations of the work presented in this thesis is presented, followed by summarising directions for future research.

7.1 Overview of research presented and novel contributions to the literature

Fatigue is an important and prevalent symptom for pwMS. However, current understanding of its experience in everyday life, as well as aetiological and exacerbating factors, is lacking. Cortisol has regulatory roles in both energy metabolism and inflammation (Sapolsky et al., 2000), so is a candidate for research exploring mechanisms underlying MS fatigue. Research into HPA axis activity in MS is limited. The empirical chapters presented in this thesis were designed to provide new information towards these research gaps, using data from a single case-control clinical study utilising an ecological momentary assessment (EMA) protocol with a 6-month follow-up. The research programme set out to examine the following broad research questions: (1) How is RRMS fatigue experienced in everyday life? (2) Is CSA elevated in RRMS in everyday life? (3) Is there evidence that CSA in everyday life is associated with fatigue across populations? (4) Is CSA associated with fatigue in RRMS?

While the majority of MS fatigue investigations have used retrospective summary measures implicitly assuming symptom constancy over time, chapter 3 directly explored the diurnal fatigue pattern by examining growth polynomials and within-subjects contextual effects in multilevel modelling (MLM); data demonstrated a substantial 39.7% of fatigue variability was attributable to within-subjects variance. Chapter 3 substantiated what

qualitative studies and anecdotal reports have suggested for some time: that MS fatigue generally gets worse throughout the day and peaks towards the late-afternoon. While the expected linear trajectory was characteristic of healthy individuals' fatigue patterns, a decelerating fatigue trajectory peaking in the late afternoon was typical for pwRRMS. Daily life stress was positively associated with fatigue in both groups, but only stressors potentially hindering goal achievement (the *Stressor Hindrance* scale) were related to fatigue.

Chapter 3 also explored the accuracy of retrospective end-of-day fatigue reports compared with daily average momentary fatigue (MF) ratings, finding that end-of-day recalled fatigue ratings generally overestimated "actual" fatigue experience in the RRMS group. This overestimation in recall was against the expectation that pwRRMS would provide more accurate recalls due to greater experience with the symptom and its reporting. Although differences between end-of-day and average MF ratings unlikely represented clinically-important differences, the study was the first to demonstrate peak and end heuristic biases in end-of-day fatigue reports in pwMS.

Chapter 4 presented the first study employing electronic platforms, compliance checks, and MLM to examine CSA in RRMS. The majority of studies of CSA in everyday life have suggested HPA axis hyperactivity in RRMS is attributable to the effects of comorbid depression (Gold et al., 2010; Gold et al., 2011; Kern et al., 2011). However, the present study demonstrated greater cortisol output during the CAR period in an RRMS sample without major depressive disorder (MDD) comorbidity or high levels of depressive symptoms. This finding supported recently published evidence demonstrating elevated CAR AUCg in an RRMS group with relatively low depressive symptoms (Kern et al., 2013). While some recent studies have examined basal CSA in RRMS, Chapter 4 presented a unique investigation of cortisol reactivity to daily stressors in RRMS. This analysis found similar cortisol reactivity to stressors, mediated by self-reported mood, in both groups, indicating the HPA axis-related stress response system is fully functional in RRMS.

A systematic review was presented in Chapter 5 of studies examining the relationship between fatigue (and chronic fatigue syndrome, CFS) and CSA measured in everyday life. Previous systematic reviews compared studies employing a variety of methods to measure CSA, including urinal, blood serum, and saliva sampling, but had only focused on CFS (Cleare, 2003; Papadopoulos

& Cleare, 2012; Tak et al., 2011). As urinal measures provide only 24h summary reports, and serum measures lack ecological validity, Chapter 5 sought to synthesise only everyday life salivary cortisol studies examining its relationship with fatigue across all populations: CFS, clinical, and nonclinical populations. The review concluded that attenuated diurnal cortisol variability (smaller CAR AUCi, flatter DCS) was associated with greater fatigue, but found little evidence to support assertions of previous reviews that CFS, or indeed fatigue generally, is associated with reduced cortisol output.

The final empirical chapter (Chapter 6) demonstrated an association between CSA in everyday life and fatigue severity in RRMS; an association which was not found in a previous study (Gold et al., 2011). The relationship between CSA and fatigue in the RRMS group was in an unexpected direction, as larger CAR AUCi (greater diurnal variability) and lower S1 cortisol were associated with higher baseline FS scores, and present in those with "clinically meaningful fatigue". While these findings appeared to contradict the conclusion of the systematic review, the only other study included in the systematic review showing a larger dynamic cortisol response to awakening associated with fatigue was in people with rheumatoid arthritis: like MS, an autoimmune disease with inflammatory pathology. The findings in Chapter 6 may demonstrate a different mechanism underlying MS fatigue than fatigue in other populations, with pro-inflammatory cytokines a possible candidate for this mechanism. Although there is some research to suggest pro-inflammatory cytokines are associated with fatigue in MS (Flachenecker et al., 2004; Gold et al., 2011; Heesen et al., 2006) more research is required.

Chapter 6 also presented the first study examining whether the CAR can predict upcoming-day fatigue in a clinical population. Previously, this hypothesis had only been examined in a population of older individuals (>50 years old) with mixed health-statuses (Adam et al., 2006). However, while there was tentative evidence (p = .054) that greater CAR AUCg was positively associated with upcoming-day fatigue in the healthy control group there was no evidence that the CAR was predictive of upcoming-day fatigue in RRMS. The longitudinal analysis examining whether CSA would predict future change in fatigue severity was also unique within a clinical population. Kumari et al. (2009) had previously found low S1 cortisol and flatter DCS predicted change in fatigue 2-3 years later in a large community sample. Although the present

study found no evidence that CSA predicted change in fatigue 6 months later, this conclusion should be interpreted with caution due to sample attrition attributable to drop-out and the development of comorbidities.

7.2 Implications

Researchers and clinicians discount important information about MS fatigue experience when using retrospective summary reports. Employing methods enabling the prospective observation of fatigue experience, as it occurs in real-time, may be valuable to both patients and observers. Beneficial effects of treatments may potentially be masked by summary reports; for example, a therapy may shorten periods of intense symptom-severity but this may not be identified by summary report.

Implementing EMA methods may provide part of the solution, but outcome measures based on EMA data still need to be developed and validated. Stone et al. (2012) detailed how EMA data could be used to derive several outcome measures. While the obvious outcome would be to compute the mean, others such as symptom variability (*SD*), symptom peaks (90th percentile), proportion-based (e.g., percentage above the midway-point on the scale), or contingent-based (e.g., morning versus evening) are possible and may be useful outcomes. Chapter 3 demonstrated that symptom experience in RRMS cannot be fully appreciated using symptom means or summaries. Stone et al. (2012) state there is sparse information available about which features of symptom experience are most important to patients. If it was found that pwMS attributed extreme peaks in fatigue severity to having the greatest impact on their daily lives, researchers and clinicians may wish to target lowering the individual's 90th percentile rating using treatments or therapies. Therefore, an important next step could be to perform a qualitative exploration investigating which facets of fatigue experience have the greatest impact on daily living.

A second problem with retrospective fatigue measures is their accuracy in reflecting actual experience, and researchers and clinicians should be aware that summary measures likely also include small but significant over-representations of most-intense and most-recent fatigue severity. While the overestimations detected in end-of-day fatigue measures may not be clinically relevant, longer recall periods may elicit greater overestimation (Broderick et al., 2008).

The findings suggested MS fatigue is likely driven by physiological disease-mediated primary mechanisms, yet fatigue experience at any given time is influenced by psychosocial factors such as stress in a very similar way to healthy individuals. These results support a model of fatigue that incorporates both primary and secondary sources of MS fatigue (Kos et al., 2008; Rosenberg & Shafor, 2005). The role of cognition and behaviour was not directly examined in this thesis, but a cognitive behavioural model of fatigue states fatigue is initiated by disease-driven triggers but maintained in a "vicious cycle" by maladaptive cognitions and behaviours (van Kessel & Moss-Morris, 2006). A feature of the cognitive behavioural model of fatigue is the rest and avoidance behavioural response to fatigue employed to alleviate fatigue by reducing demands but is actually associated with increasing fatigue levels (Skerrett & Moss-Morris, 2006), potentially by disturbing sleep-wake cycles (Strober & Arnett, 2005). The findings in Chapter 3 suggest only when demands are perceived as hindering goals will they manifest as elevated fatigue in RRMS, and this information could be used to support messages to patients explaining that avoidance of all activity is maladaptive. Future research may wish to examine goal disturbance as a potentially modifiable factor that could be incorporated into cognitive behavioural or stress management therapies for MS fatigue. The abandonment of unrealistic goals and the reformulation of new, realistic, goals may result in fewer stressors being interpreted as hindering goal achievement, and potentially alleviate the daily life stress-fatigue relationship.

Greater cortisol output in RRMS did not appear to be caused by potentially-modifiable psychosocial causes such as chronic stress or depression, which implies it is likely caused by a disease mechanism, with heightened pro-inflammatory cytokines a probable causal mechanism (Ysrraelit et al., 2008). MS fatigue was actually present in individuals with CARs resembling that of healthy individuals; heightened CAR output in RRMS seemed an adaptive response to the condition. This finding has limited immediate benefit to pwMS. However, it does suggest a testable causal mechanism for MS fatigue: low waking cortisol levels inhibiting appropriate regulatory control over the heightened inflammation in MS, followed by a large compensatory awakening response (CAR AUCi). This hypothesis would suggest MS fatigue manifests in situations where inflammation is high and regulatory control from the HPA axis is inadequate. It would be difficult to test this hypothesis in

everyday life due to the unreliability of saliva samples (low correlation between saliva and serum levels, $r < .30$) for inflammatory biomarkers (Cox, Pyne, Gleson, & Callister, 2008; Fernandez-Botran, Miller, Burns, & Newton, 2011); however, a lab-based study taking serial measures of serum would be appropriate. A greater understanding of the aetiology of MS fatigue is needed.

7.3 Strengths and limitations

The empirical work presented in this thesis had several methodological strengths but, equally, there were some limitations to highlight. The method described in Chapter 2 sought to address several limitations present in previous MS research including a lack of ecological validity, low number of assessments and assessment-days in EMA designs, and lack of compliance control in EMA studies. The over-arching strength of the research presented in this thesis was its focus on prospective real-world data yielding findings with high ecological validity representing everyday life experience for pwRRMS.

7.3.1 Generalisability

One of the potential limitations of the research presented was the focus on a single sample of participants across all empirical chapters. While this strategy enabled a comprehensive examination of phenomena, the generalisability of the sample is an important consideration.

7.3.1.1 Eligibility criteria and sample homogeneity

The eligibility criteria were justified *a priori* in section 2.2.1 of Chapter 2. Firstly, only those individuals with an RRMS diagnosis were eligible for the clinical group. The homogeneity of the sample, while limiting the generalisability of findings to only those with RRMS, was considered a methodological strength. At several points within the thesis, the importance of recruiting a relatively homogeneous MS sample to enable appropriate interpretations of findings was emphasised. As described in Chapter 1, pwMS have very different disease courses; the three broad MS-types (RRMS, SPMS, and PPMS) each have different prognoses, speeds of functional degradation, levels of disability, and pathological mechanisms (Lassmann, van Horssen, & Mahad, 2012; Rice, Cottrell, Wilkins, & Scolding, 2013). For example, HPA axis activity and inflammatory cytokine levels appear to be elevated in PPMS

compared to SPMS, and in SPMS compared to RRMS (Ysrraelit et al., 2008). Therefore, rather than a vague understanding of phenomena examined across all MS-types, the findings provided a focused examination of RRMS. Ideally, replication of the current study would recruit large sub-samples of participants from each MS-type; however, a protocol requiring fine motor-skills and sound cognitive functioning to operate both the handheld and salivettes would likely present difficulties to progressive MS populations. Extensive piloting and probable protocol-modifications would be required.

The eligibility criterion that no individual within the RRMS group was within 3 months of relapse or corticosteroid treatment ensured all pwRRMS taking part were clinically stable. This was an important methodological strength in terms of CSA, as urinary cortisol levels, as well as pro-inflammatory cytokines have been elevated in the relapse-phase (Ysrraelit et al., 2008), which is characterised by neural inflammation. Clearly, exogenous corticosteroid use would also affect endogenous CSA. Previous studies of salivary cortisol in RRMS have recruited only individuals in remission-phase and outside of corticosteroid treatment (Gold et al., 2010; Gold et al., 2011; Kern et al., 2013; Kern et al., 2011), so this criterion facilitated study comparisons.

Previous studies have revealed the extent of comorbidity in MS. Marrie et al. (2008) found 77.1% of 8983 pwMS self-reported at least one physical comorbidity; although a more conservative estimate of 36.7% was reached by including only those conditions defined as "very likely" accurately self-reported (e.g., cancer and diabetes). Psychiatric disorders are also prevalent in MS with life-time prevalence of MDD at around 50% (Goldman Consensus Group, 2005) which is far higher than the 16.2% rate found in the United States general population (*n* = 9090) (Kessler et al., 2003). By eliminating comorbidities from the RRMS sample, it was possible to be more confident that effects were attributable to RRMS, but the findings may not be generalisable to the RRMS population at large because such a substantial proportion have other conditions that may affect cortisol and/or fatigue.

Most pwRRMS are taking some form of DMT and/or symptom-relieving medications. In the present study, 61% of participants in the RRMS group were using a DMT. Individuals taking DMTs were included so findings were more generalisable; however, it should be acknowledged that while DMT-use was

incorporated into statistical analyses, it was possible these medications elicited effects on fatigue and cortisol levels. There was no evidence to suggest this was the case, but it was not possible to check each type of DMT separately given low numbers.

7.3.1.2 Recruited sample

Sampling bias (those taking part in the study systematically differing from those who did not) is a difficulty within all research with volunteer participants, given the individual's ethical right to refuse participation. It was probable that those who did take part were more motivated to participate than those who did not. The recruitment procedure made it often unfeasible or impossible to collect descriptive data on those who did not agree to take part; in clinic-based recruitment, individuals were able to refuse participation even before meeting with the researcher. Not having such information about study-refusals is a limitation as this information would enable sampling bias to be directly assessed. Therefore, it was important to assess the representativeness of the recruited sample to identify any signs of sampling bias, particularly given EMA protocols can impart a relatively high burden on the participant.

7.3.1.2.1 Stress and Fatigue

Feasible examples where burden might be a particularly problematic would include when an individual does not want to add additional tasks to a high-stress lifestyle, or not wanting to take part while feeling particularly fatigued. Both these examples would be problematic for the present research given its focus on stress and fatigue. However, there was little evidence of either being a problem here. Average self-reported stress measured by the PSS in the RRMS group at baseline (M = 18.6, ±6.0) was 25% higher than in a study providing PSS normative scores for pwMS (n = 446, M = 14.9, ±8.0) (Wu & Amtmann, 2013). While this refutes the possibility that people experiencing high stress were less likely to take part, it may suggest that individuals with stressful lifestyles were attracted to take part in order to contribute to a study examining its effects on MS. While the average PSS in the control group (M = 13.2 ±6.5) was somewhat lower than most-recent normative scores (n = 2000, females M = 16.1 ±7.6, males M = 15.5 ±7.4) (S. Cohen & Janicki-Deverts, 2012), it should be noted these normative scores were obtained from a general population which would contain individuals with different health

conditions, and normative scores from the PSS during development (*n* = 2270) were *M* = 13.7 ±6.6 for women and *M* = 12.1 ±5.9 for men (S. Cohen et al., 1983).

While many studies have reported greater fatigue in progressive MS-types compared to RRMS (Flachenecker et al., 2002; Kroencke et al., 2000; Patrick et al., 2009), fatigue is typically reported as affecting 60-80% of heterogeneous MS groups (Hadjimichael et al., 2008; Lerdal et al., 2003; Minden et al., 2006). In the present study, 55% of the RRMS group had "clinically meaningful fatigue" which, while slightly below the lower end of the prevalence range for fatigue in all MS-types, was similar to the 58% rate (scoring ≥5 on the FSS) observed in a RRMS sample elsewhere (Bakshi et al., 2000).

7.3.1.2.2 Mental health

Depressive symptoms and anxiety were also examined. In a large study of pwMS (*n* = 4178), HADS normative scores were obtained for both the HADS-D (*M* = 8.9 ±4.4) and HADS-A (*M* = 8.2 ±4.3) subscales (Jones et al., 2012). In the present study, participants were excluded based on HADS-D scores ≥8, so although the level of depressive symptoms was far lower in the present study (HADS-D *M* = 4.0 ±2.3) this was unsurprising but still important in terms of representativeness. HADS-A scores for the present study (*M* = 7.5 ±3.9) were broadly similar to the normative score. Compared to data from a general population (*n* = 1792) (Crawford, Henry, Crombie, & Taylor, 2001), HADS-D scores for the control group (*M* = 2.1 ±2.3) were somewhat lower than normative scores (*M* = 3.7 ±3.1) and, similarly, HADS-A scores were lower (*M* = 4.8 ±3.1) than normative scores (*M* = 6.1 ±3.8). Again, these differences should be considered in light of the general population normative scores not being based only on healthy individuals. In summary, there appeared no reason to suggest individuals rejected participation based on heightened stress-levels, fatigue, or depressive or anxious symptoms. However, one cannot completely rule out that self-selection bias was present in the sample of volunteers.

7.3.1.2.3 Other variables establishing representativeness

Randomised recruitment from multiple sites throughout the UK of individuals meeting eligibility criteria would be the ideal strategy for obtaining a representative sample. However, this was not possible for logistic and

economic reasons. This is a limitation, although all previous studies of salivary cortisol in MS have recruited from a single clinical site and relied on volunteers. The present study was able to recruit from two NHS outpatient centres: the General Hospital in Southampton, and Guy's and St Thomas' Hospital in London. The "catchment areas" for the two clinics were quite diverse, particularly in Southampton where patients were referred from urban (Southampton, Bournemouth, and Basingstoke) and rural (New Forest, and rural Dorset) areas. While patients attending the London clinic were generally living in the highly-urbanised Greater London area, some had been referred there from smaller towns in Surrey and Kent. While the study succeeded in recruiting some pwRRMS from non-NHS sources, this was only a small number as most individuals contacting the researcher via such sources were ineligible.

The gender ratio of the RRMS sample (76% female) was representative compared to a recent study reporting that women accounted for 72% of MS prevalence in the UK in 2010 (Mackenzie et al., 2014) based on the General Practice Research Database (Walley & Mantgani, 1997): a longitudinal database (1990 – 2010) containing patient details from a representative sample of UK general practices. A broad age-range (27-56 years) of participants were also recruited to the RRMS group. According to the available data from the epidemiological study by Mackenzie et al. (2014), around 70% of pwMS in the UK are between 20 and 60 years of age. A substantial proportion of over-60s have progressed to SPMS from RRMS given typical onset at around 30 years (Richards et al., 2002) and usual progression to SPMS within 15 years for 65% of pwRRMS (Koch et al., 2008).

The education level of the sample in both groups was relatively high compared to national averages. The Office for National Statistics (2013) stated that 38% of individuals over 21 years and of working age were educated to graduate-level in the UK, whereas both the RRMS (47%) and control (71%) groups reported higher rates of graduate-level education. The UK MS Register (an online database collecting ongoing self-report data) provided data suggesting only 5.0% of 5406 pwMS of working age (63% RRMS) were in full-time employment, and 36.9% in part-time employment (Ford et al., 2012). Employment within the RRMS group in the present study (52.6% full-time, 26.3% part-time, 21.1% unemployed) was much higher, and would suggest less disability within the sample than which is typical. This should be expected

given the eligibility criterion requiring individuals be relatively mobile; greater mobility would likely enable employment. However, self-reported EDSS scores did not reflect this, with a mean score of 4.35 suggesting not-insignificant disability levels within the sample. It was possible that those most able to assimilate the EMA protocol into their everyday lives were in jobs either based in an office or with high levels of autonomy. Speculatively, these jobs may have required a degree to obtain (explaining the high education level) and may be more amenable to high levels of disability than, for example, manual jobs.

7.3.2 Extraneous causes of fatigue

While the study design aimed to eliminate or control for factors that may account for fatigue variance, there were some which were not possible to account for, or overlooked. It would have been useful to collect data around daily working patterns while undergoing the 4-day EMA protocol. Although momentary assessments provided information about "Working/Studying" with any given event, this did not provide accurate information about the length of time in work or indeed whether the individual was working from home or in the workplace. This information would have helped to elucidate the impact of work on fatigue experience.

Iron-deficiency anaemia is associated with fatigue (Cook, 2005). Previous research has suggested prevalence of iron-deficiency anaemia between 2.5% and 3.5% in pwMS (Horton et al., 2010; Kang et al., 2010) compared to a prevalence of 0.6% in the general population (Kang et al., 2010). A strength of the study was that the eligibility criteria excluded individuals with anaemia, and both recruitment centres confirmed patients were tested for iron deficiency in routine blood tests. However, iron supplements have been shown to improve unexplained fatigue even in non-anaemic iron-deficient women (Vaucher et al., 2012; Verdon et al., 2003) and therefore serum iron levels may have been an important covariate which was not controlled.

Finally, MS fatigue has been described as heat-sensitive, and affected by sunlight and humidity (Stuifbergen & Rogers, 1997), but it was not possible to objectively observe any of these variables within the study. It may also have been informative to incorporate motion detectors, such as accelerometers (Yang & Hsu, 2010), into the study to objectively observe physical activity levels as a more accurate covariate than self-reported activity.

7.3.3 Momentary assessment

A major strength of this study was the assessment design within the EMA protocol, which required assessments at strategic time-points while also taking steps to maximise compliance. It might be assumed that repeated assessment of fatigue would encourage greater symptom-focussing (measurement reactivity). While the study was not designed to assess the presence of measurement reactivity, previous evidence from EMA studies (discussed in detail in section 2.1.4 of Chapter 2, p. 43) has suggested its effects are minimal (for example, Sonnenschein et al., 2006; Stone et al., 2003).

The variable-occasion design (A1–A6 events) chosen to measure daily fatigue experience enabled a representative sample of fatigue experience. While it is clear that six assessments per day cannot observe all variability in either fatigue or cortisol over the 1000h to 2000h period, the variable-occasion design meant there was no reason to expect any systematic bias. The assessment of fatigue within days was restricted to between 1000h and 2000h. While this prevented sleep-wake processes from confounding the diurnal fatigue pattern, a measure of fatigue upon awakening would have been informative, even if examined separately from the rest of the diurnal pattern. While the original idea was to incorporate an all-encompassing single-item MS fatigue measure, measuring mental and physical fatigue separately in respective items would have enabled examination of whether diurnal fatigue patterns are different for mental and physical fatigue and whether stress and mood is differentially associated with these fatigue dimensions.

It has previously been suggested that single-item scales, such as that used to measure MF, are more likely to contain ceiling effects than longer scales (McHorney, 2000). As such, it could be argued that the drawn-out peak in fatigue observed in the RRMS group was simply an artefact of such a ceiling effect. However, this was considered unlikely given the maximum score on the 0-10 VAS was used sparingly in both groups (< 3% in the RRMS group, < 1% in the control group).

7.3.3.1 Salivary cortisol assessment protocol

The salivary cortisol protocol had several strengths to highlight. The CAR was measured over 4 consecutive weekdays, which was 2 days more than any

previous study of everyday life CSA in MS (Gold et al., 2010; Gold et al., 2011; Kern et al., 2013; Kern et al., 2011). Consecutive days meant participants could not self-select assessment days to suit schedules; while the Gold research group's studies measured cortisol over 2 consecutive days, the Kern research group's studies measured cortisol over 2 days within a 1-week or 2-week period. No previous study of CSA in MS reported controlling for weekday-weekend effects in CSA (Kunz-Ebrecht et al., 2004; Schlotz et al., 2004), or controlling for important covariates such as smoking or exercise (Adam & Kumari, 2009) or behaviours such as brushing teeth before sampling (Kudielka et al., 2012). All these factors had the potential to introduce bias to previous findings.

Many studies have now highlighted the importance of delays in S1 cortisol samples confounding CAR measurement (Dockray et al., 2008; Griefahn & Robens, 2011; Okun et al., 2010; N. Smyth et al., 2013) and it was a limitation in the present study that awakening time was not objectively observed. Although awakening time was time-stamped as the time of S1 event alarm or manual initiation, if the participant did wake before their pre-set S1 alarm then there was no means of detecting whether immediate manual initiation occurred or not. This limitation is far from unique to the present study; no study included in the systematic review in Chapter 5 or any of the studies exploring salivary cortisol outcomes in RRMS (Gold et al., 2010; Gold et al., 2011; Kern et al., 2013; Kern et al., 2011) objectively observed awakening. Allowing participants to use the S1 event alarm as an auxiliary alarm clock went some way to alleviating this limitation, although it was not the "gold standard" of objective observation of awakening.

A potential limitation was the fact cortisol reactivity to stress was examined using cortisol samples provided simultaneous to self-report stress measures. A previous study has suggested lagged-associations would be more appropriate, as salivary cortisol responses lag behind psychological responses to stressors by 15 to 20 minutes (Schlotz et al., 2008). With daily life stress operationalized "Since the last event…" to capture all stressors throughout the day, this somewhat resolved the lagged effects issue. However, ideally, the saliva sample would have been provided 15 minutes after the mood measure for testing the mediating role of distress in the stress-cortisol relationship.

7.3.3.2 Compliance

A previous study found that participants unaware of being monitored for compliance in a saliva sampling protocol in everyday life achieved an 80% compliance rate in a clinical group (fibromyalgia) and only 62% in a healthy control group (Broderick et al., 2004). Previous studies with salivary cortisol measurement in MS have not reported compliance rates, and did not take steps to maximise compliance. Compliance rates in the present study were excellent: 96.6% of fixed-occasion (S1–S3 events) samples, 89.7% of variable-occasion (A1-A6 events) samples, and 91.1% of variable-occasion momentary self-reports were provided. The present study included many features to promote compliance that previous studies in MS have lacked: electronic prompting, electronic time-stamps, and accurate transfer of random codes to salivette label. Use of an objective measure of cortisol sampling compliance such as the Medication Event Monitoring System (MEMS), which time-stamps the opening of container caps, would have further increased confidence that good levels of compliance was achieved. The self-report compliance rate was far higher than the 64% average compliance rate observed by the only other study collecting MF assessments in MS in everyday life (E. Kim et al., 2010).

7.3.4 Follow-up

The inclusion of a follow-up measurement of the fatigue outcome in Chapter 6 was a strength, as no previous clinical study had examined the relationship between CSA and fatigue with a longitudinal perspective with the potential to make causal inferences. However, while the 78% response-rate was reasonable, re-application of the eligibility criteria led to only 65% of the original sample retained in the follow-up. This was primarily due to reporting the diagnosis of comorbidity, clinical relapse, or becoming pregnant making responders ineligible. Future research incorporating strict eligibility criteria and follow-ups should be aware of the likelihood that a substantial proportion of respondents may become ineligible over time.

The longitudinal analysis was also inhibited by insufficient variability in fatigue from baseline to follow-up: only four individuals made changes in fatigue statuses that were “clinically meaningful” based on the FS. It would be beneficial in future studies to have multiple follow-ups from which to model change in fatigue severity.

7.4 Future research

Future research should seek to address those methodological limitations indicated in the previous section, as well as retain some of the strengths highlighted. Several directions for future research have been suggested throughout this chapter. Some other ideas for future research are detailed below.

7.4.1 Symptom recall

Both research and clinical practice would benefit from a greater understanding of how fatigue reports are provided by pwMS. Future research should examine the accuracy of fatigue reports provided over periods longer than 1 day. For example, recall of fatigue (item: "worn out") in rheumatology patients was less accurate with increasing number of days' recalled (1-day, 3-day, 7-day, and 28-day) while other measures of fatigue ("energy", "tired", "full of life") were least accurate in 7-day recalls (Broderick et al., 2008). Future research in MS may also wish to examine whether the peak and end effects observed in 1-day recall are also present in larger recall periods.

In a previous cross-sectional study, the tendency to attribute somatic symptoms to MS was associated with greater mental fatigue in a group of pwMS (Skerrett & Moss-Morris, 2006). Those attributing somatic symptoms to physical illness tend to focus more on the symptom and have negative interpretations of the symptom. Future studies could be designed to see whether such tendencies manifest as over-representation of symptoms in recall.

7.4.2 Stress reactivity

While the current data demonstrated cortisol reactivity to daily stressors in RRMS that was similar to that observed in healthy individuals, the literature would still benefit from a study examining cortisol reactivity to a experimentally-induced psychosocial stressor in MS. While such a study has been performed twice previously (Ackerman et al., 1996; Heesen, Schulz, et al., 2002) both studies' results were confounded by the effects of the CAR when conducting the study in the early morning. Future studies should choose a time later in the day, and use a well-standardised psychosocial stressor, such

as the Trier Social Stress Test (Kirschbaum, Pirke, & Hellhammer, 1993) to enable comparisons with studies in other populations. The two previous studies by Ackerman et al. (1996) and Heesen, Schulz, et al. (2002) demonstrated inadequate cortisol responses to an acute stressor in pwMS. Inadequate stress regulation via the HPA axis might contribute to the manifestations of symptoms and, potentially, explain the increased relapse risk associated with stressful life events (Artemiadis et al., 2011; Mohr et al., 2004). Such a study would have ethical challenges, with careful elucidation of the different mechanisms underlying stressful life events and short-term acute stressors required to ensure participants were not put at risk of relapse.

Future research may also seek to prospectively observe CSA, symptoms, and relapses over the longer-term in RRMS, with CSA measured over several days every month for 1 or 2 years, alongside concurrent stressful life event, symptom, and relapse reports. Such a study would help to elucidate the role of HPA axis-related stress regulation in the manifestation of symptom exacerbations/relapses in RRMS. Given that RRMS relapse-phase is accompanied by increases in inflammation (Ysrraelit et al., 2008), it seems paradoxical that stress (and associated increases in cortisol output) is associated with an increased risk of a relapse. One proposed hypothesis is that individuals falsely attribute relapses to stressful life events, and it is actually when the stressful life event ends that cortisol levels are reduced, inflammation overshoots, and relapse occurs (Mohr & Pelletier, 2006). This is a hypothesis that could be tested with an extended longitudinal design implementing EMA.

7.4.1 Progressive MS types

While this thesis has frequently stated the importance of examining phenomena in a homogeneous MS sample, this resulted in findings which may not be generalizable to individuals with SPMS or PPMS. Research focussing on progressive MS types is required, in terms of all the research questions examined in this thesis. Diurnal fatigue patterns in SPMS and PPMS have not yet been examined, while future research may wish to examine the relative importance of HPA axis activity and psychosocial factors to within- and between-subject variance in everyday life. Researchers would need to comprehensively pilot electronic platforms and the feasibility of saliva sampling in a population of individuals with more advanced MS who likely have

increased limb apraxia (inability to perform coordinated movements) (Kamm et al., 2012) and greater cognitive dysfunction (Bobholz & Rao, 2003) than the population observed in this thesis.

7.5 Conclusions

The primary aims of the research presented in this thesis were to provide a detailed investigation of (1) fatigue experience within-individuals and within-days; (2) the circadian rhythm of cortisol in RRMS; and (3) associations between RRMS fatigue and cortisol both at baseline and longitudinally. The analysis in Chapter 3 confirmed the importance of within-subjects variability in fatigue experience in RRMS. Fatigue in RRMS followed a typical increasing (but decelerating) diurnal pattern different to a "normal" linear pattern in healthy individuals. Threatening stressors than potentially hinder goal achievement were identified as psychosocial contextual factors that exacerbate MS fatigue, although this contextual effect was also evident in the control group. At the point of experience, fatigue was associated with distress (positive association) and positive mood (negative association); lagged analysis was unable to identify causation in either direction.

The analysis in Chapter 4 provided evidence of HPA axis hyperactivity (greater CAR output) in everyday life in RRMS, without comorbid MDD. There was no evidence of a dysfunctional HPA axis-related stress response system in RRMS. Chapter 6 demonstrated certain facets of the CAR (low S1 cortisol, high AUCi responses) were associated with fatigue in RRMS, which differed from the conclusions of the systematic review in Chapter 5. The systematic review suggested attenuated diurnal variability in cortisol was most relevant to fatigue in CFS and other clinical and nonclinical populations. This contradiction suggests fatigue in RRMS is likely caused by different mechanisms than in CFS or other clinical populations. Future research should seek to extensively explore the role of pro-inflammatory cytokines in mediating the cortisol-fatigue relationship in RRMS.

Appendices

Appendix A – **Cohen's** d formula

To facilitate thorough interpretation of and comparisons with previous research, Cohen's *d* effect sizes (J. Cohen, 1988) were computed and presented based on independent or paired samples t-tests using aggregated data. Cohen's d is presented where sufficient data was available for its computation. Cohen's *d* is appropriate to present sizes of effects for outcomes that are normally distributed and is used as an indicator of effect size for those that are not normally distributed. Interpretation of Cohen's *d* for nonparametric data can be problematic (McGough & Faraone, 2009); however, using a consistent measure of effect size throughout simplified comparisons.

The following formulae were used, with pooled standard deviation computed differently in independent and paired samples:

Equation

$$Cohen's\ d = \frac{M_1 - M_2}{SD_{pooled}}$$

For independent samples:

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + (n_2 - 2)}}$$

For paired samples:

$$SD_{pooled} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}$$

Where M_i is the mean for group i, SD_i is the standard deviation for group i, and n_i is the sample size for group i.

Appendix B – HPA axis pharmacological challenge tests

Dexamethasone is a corticosteroid which mimics the actions of cortisol by binding with glucocorticoid receptors at the pituitary, inhibiting the secretion of ACTH and then cortisol. Dexamethasone passes the blood-brain barrier poorly (Meijer et al., 1998) so has minimal inhibition on the hypothalamus and CRH production. The lack of circulating cortisol resulting from dexamethasone prompts a reactionary increase in CRH and arginine vasopressin (AVP) production.

Feedback sensitivity of the HPA axis is dependent on the number of glucocorticoid receptors and their binding affinity, and can be examined by probing the response to CRH under conditions of minimal regulatory feedback from cortisol (Heuser et al., 1994; von Bardeleben & Holsboer, 1989). Hyper-response to the Dex/CRH test is believed to indicate decreased glucocorticoid receptor capacity, resulting in the decreased inhibition of CRH acting on the pituitary gland. The Dex/CRH test is frequently used in diagnosing Cushing's disease (Yanovski, Cutler, Chrousos, & Nieman, 1998) and a hyper-response is indicative of depression (Heuser et al., 1994) and seems to be normalized by antidepressant medications (Ising et al., 2007).

Described below are the standard protocol and rationale behind the dexamethasone suppression test (DST) (Liddle, 1960) and the dexamethasone/CRH (Dex/CRH) test (Heuser et al., 1994). Recent research has tended to move away from using the DST due to the Dex/CRH test being more sensitive to detecting HPA axis dysfunction.

Dexamethasone suppression test

In the DST, 1.0 mg of dexamethasone is administered at 2300h, with a single blood sample for cortisol level taken the following morning at 0800h. Cortisol suppression (< 50 nmol/L) indicates a properly-functioning HPA axis.

Dexamethasone/CRH test

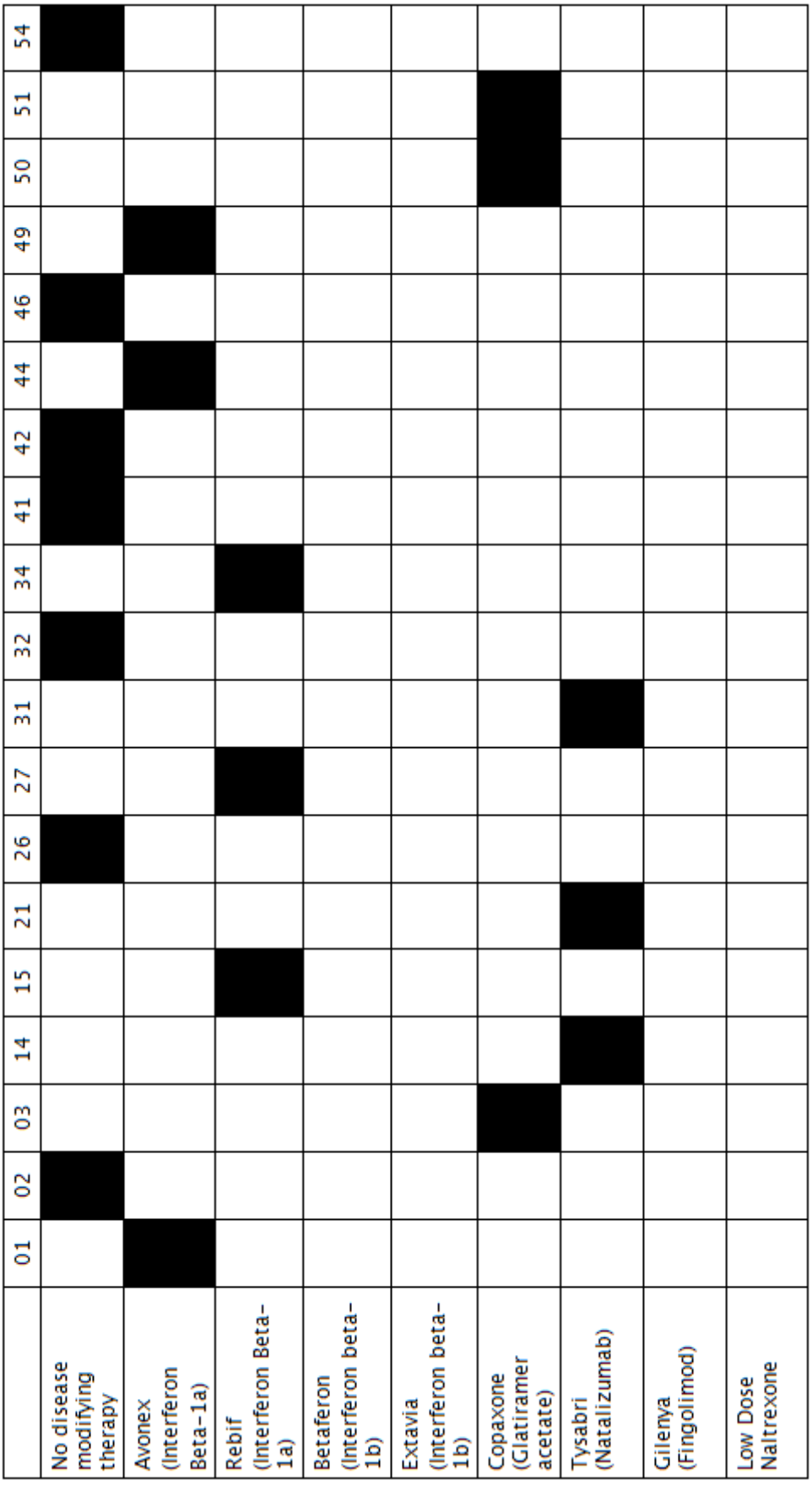
In the standard Dex/CRH test, 1.5mg of dexamethasone is administered orally at 2300h the day prior to testing. Testing then involves serial blood samples at 1500h, 1530h, 1545h, 1600h, and 1615h, with 100 µg of human

CRH administered intravenously within 30 seconds of the 1500h. The patient or participant rests supine (lying down face up) for the test.

Plasma ACTH and cortisol at 1500h therefore reflect the suppressive effects of dexamethasone, with suppression (< 40 ng/ml) indicative of a properly-functioning HPA axis. The CRH injected shortly afterwards, combined with the increased endogenous CRH and AVP (which synergises the effect of CRH) overrides the dexamethasone in non-suppressors and a response is observed. Thus, the later samples reflect the additional effect of CRH stimulation. Researchers may compute markers such as ACTH/cortisol AUC or $\Delta max_{cortisol}$ (subtracting 1500h level from peak level) to assess response to CRH.

Appendix C – RRMS group medication regimens

Participant numbers with medication regimen. Black shading indicates receiving this disease modifying therapy (DMT) at baseline.



Participant numbers with medication regimen. Black shading indicates receiving this disease modifying therapy (DMT) at baseline.

	55	62	63	66	68	69	70	71	72	73	75	79	80	82	83	84	86	87	88
No disease modifying therapy		■	■		■		■			■			■			■	■		
Avonex (Interferon Beta-1a)	■											■							■
Rebif (Interferon Beta-1a)				■							■			■					
Betaferon (Interferon beta-1b)																			
Extavia (Interferon beta-1b)																			
Copaxone (Glatiramer acetate)						■		■						■				■	
Tysabri (Natalizumab)									■						■				
Gilenya (Fingolimod)																			
Low Dose Naltrexone	■																		

Other treatments

Type of symptom treatment and name of drug stated. Participant taking drug reported in brackets.

Constipation

- Movicol (54)

Erectile dysfunction

- Viagra (71)

Insomnia

- Zoplicone (49)

Muscle antispasmodic

- Baclofen (26, 49, 82)
- Diazepam (49)
- Tinazidine (3)

Neuropathic pain

- Gabapentin (26, 49, 55, 72, 79)

Other

- Aspirin (14)

Urinary urgency, incontinence, or infection

- Detrunorm (15)
- Nitrofurantoin (51)
- Sanctura (55)
- Solifenacin (51)

Vitamin/Mineral deficiency

- Calchew (46)

Appendix D – Participant information sheet

UNIVERSITY OF Southampton
School of Psychology

Daily Hassles and Cortisol in People with and without Multiple Sclerosis

Participant Information Sheet

Research Team: Mr Daniel Powell (co-ordinator), Dr Wolff Schlotz (supervisor), Professor Rona Moss-Morris (supervisor), Dr Christina Liossi (supervisor).
School Ethics number: 589
NHS Ethics number: 11/SC/0333

Thank you for your interest in this research study. Please read what is presented here carefully before deciding whether to take part. Part 1 provides information about the main details of the study and what it will involve for you. Part 2 will give additional details for you to read if the information in Part 1 has interested you and you are considering participation.
You will have the opportunity to discuss the details of this information sheet with the co-ordinator before you will be asked to give consent to taking part.

Part 1

What is the purpose of the study?

Every person experiences daily hassles to a certain extent, whether this be managing finances, the demands of children, or other lesser or more excessive demands. This study aims to provide new information about the possible relationship between daily hassles and a hormone called *cortisol* which can be found within saliva, and how these may affect certain Multiple Sclerosis (MS) symptoms. It is important to try and gain a more detailed understanding of how everyday life activities and hassles may affect the experience of having MS, and we intend to do this by comparing a group of people with relapsing-remitting MS to a group of people with no chronic or acute illness.

This study will also form part of the research requirements towards a PhD qualification for the co-ordinator, who is himself a postgraduate research student at the University of Southampton.

Why have I been chosen?

You have been invited to take part as this study is looking for people with relapsing-remitting MS who live in and around the Southampton area. A further 40 - 50 individuals with relapsing-remitting MS will be asked to take part, as well as a further 40 “healthy control” individuals in order to compare.

Do I have to take part?

It is completely up to you whether or not to take part. If you do decide you would like to take part, you will be asked to sign a consent form. You would then be given a copy of this information sheet and your signed consent form to keep and refer to if needed. Please be aware that you are free to withdraw at any time during the research without giving a reason.

What will happen to me if I take part?

You will be involved in the research for one week (most of which will be conducted while you do your usual daily activities), plus some follow-up questions six months later. The different steps involved are described in sequence below:

1. Firstly, you will be asked to come to the School of Psychology at the University of Southampton for an introductory session. Here, you will be asked to complete several questionnaires on a PC, and you will be trained in the use of Salivettes (used for collecting saliva) and a specially-programmed handheld computer (which prompts the saliva collection – see below). **This session will last approximately 1 hour**.

2. The next step will take place over **four weekdays** while you undergo your usual daily activities and will be guided by handheld computer. You will be prompted to collect saliva samples by alarm and instructions will be presented on the handheld's screen. It will also ask some questions for you to answer each time using the handheld. There are 10 alarms per day (all except the last one each day instructs to collect saliva), and **each handheld presentation should take no longer than 5 minutes.** You will be fully trained in what to expect during the introductory session, and will receive a small instruction leaflet to help you.

3. Once all four days are complete, you will be asked to return your used Salivettes and handheld computer to the School of Psychology, where you will be thanked and debriefed. **This will take no longer than 30 minutes.**

4. Six months later, you will be sent some questionnaires for you to complete, which **will take no more than 10 minutes**. You will be asked to return these in the mail (you will be provided with an addressed freepost envelope).

Expenses and Reimbursement.

Each participant will be reimbursed with a total of £40 (2 x £20) for taking part in this study. This is intended to cover the travel expenses of participants to get to and from the University, and also to compensate for the time taken to complete the study. One £20 reimbursement will be provided upon attendance at the introductory session and the final £20 when equipment and samples are returned.

What will I have to do?

You will leave the introductory session with 36 Salivettes (in four bags of nine) for the four study days, as well as the handheld computer. A Salivette is a small plastic vial containing a synthetic swab which, when chewed, collects saliva.

Participants will need to be awake by 8.30am on each day, and the last alarm will be no later than 9pm. During the first 45 minutes after waking, you will be asked to take three samples (immediately upon waking, 30 minutes later, and 45 minutes later). **During this 45 minute period, you will be asked to refrain from eating, drinking (only water permitted), smoking, brushing teeth, and exercise in order to prevent contamination of morning samples.** You will then hear a further six alarms at randomly spread intervals throughout each day which will prompt you to take a saliva sample and present some questions to be answered.

At every alarm requiring a saliva sample, a three-digit code will be presented by the handheld which you will need to write on the label of the Salivette you are using. You will then need to gently chew on the synthetic swab contained within the Salivette for approximately one minute. A final alarm will sound at 9.00pm which will ask some general questions about the day overall, but will not ask for a saliva sample to be taken.

Please note that you will not be expected to remember everything as the handheld will guide you thoroughly on each day, although you will need to keep the handheld with you at all times and follow the instructions carefully. You will be asked to store used Salivettes in a fridge (within their sealed bags) at the end of each day until the date of your appointment to return them to the School.

What are the possible risks of taking part?

There is a very small chance that this study may reveal very high cortisol levels or a high level of depressive symptoms. In either circumstance, you will be appropriately informed at the earliest opportunity and advised to make an appointment with your GP for a consultation. **We will not inform your GP or neurologist unless you ask us to.**

What are the possible benefits of taking part?

There will be no direct benefit to the participant by taking part, but the information gained from this study will help our understanding of how people with Multiple Sclerosis deal, both psychologically and physiologically, with the hassles experienced in everyday life. This information may then in some way help explain how some of the symptoms experienced come about.

What if there is a problem?

Any complaint about the way you have been dealt with during the study will be addressed. The detailed information on this is given in Part 2.

Will our taking part in the study be kept confidential?

Yes. All information about your participation in this study will be kept strictly confidential. The detailed information on this is included in Part 2.

This completes Part 1 of the Information Sheet

If you are interested and considering participation, please read the additional information in Part 2 before making any decision.

Part 2

What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the research co-ordinator who will do their best to answer your questions by calling 023 8059 4719. If you remain unhappy and wish to complain formally, you can do this by contacting Dr Martina Prude at the Research Governance Office of the University of Southampton:

Research Governance Office
George Thomas Building 37
Room 4055
University of Southampton
Highfield
Southampton SO17 1BJ
Tel: 023 8059 8848.

As outlined in Part 1, the planned procedure and sampling is considered safe. In the unlikely event that something does go wrong and you are harmed due to someone's negligence then you may have grounds for a legal action for compensation against the University of Southampton but you may have to pay your legal costs.

Will my taking part be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. As outlined in Part 1, if anything is discovered that may be useful to your GP or neurologist in your treatment, we would give you the details and ask you to arrange a consultation with them (we will contact them only at your request).

What will happen to the saliva samples I give?

Your samples will be handled in accordance with the Human Tissues Act (2008) which ensures they are kept with strict confidentiality and professionalism. They will only be identifiable by your unique participant number, which is protected by a code known only by the research team. Samples will be stored at the University of Southampton until all participants have completed their four days of sampling, at which point they will all be sent to the University of Trier, Germany (who have a specialist salivary cortisol analysis lab) where they will undergo analysis and then be appropriately destroyed. **Samples will only be used for the analysis of cortisol content.**

What will happen to the results of the research study?

The results of the study will be published in relevant academic journals and may be presented at conferences to people with an interest in MS and/or the relationship between psychosocial factors and the cortisol found in saliva. They will also be presented as part of a PhD thesis prepared by the research co-ordinator. A brief report from this study will be made available on the MS Society website (http://www.mssociety.org.uk/research) and the University of Southampton website (http://www.soton.ac.uk/psychology) at the study's conclusion.

Who is funding the research and who has reviewed the project?

The research study is supported and funded by the University's School of Psychology and the research student by the Economic and Social Research Council, and approval has been given by the School of Psychology Ethics Committee and University Research Governance Office, as well as an NHS Ethics Committee.

If you have any further questions please feel free to ask the study co-ordinator:

Daniel Powell
Tel: 023 8059 4719
Email: daniel.powell@soton.ac.uk

What do we do now?

You will be contacted by telephone by Daniel Powell sometime in the next week in order to discuss what you have read within this information sheet and to see whether or not you would be interested in taking part in the research study. If you do wish to take part, then a date and time will be scheduled to attend the introductory session.

Many thanks for taking the time to read this information sheet.

UNIVERSITY OF Southampton
School of Psychology

Daily Hassles and Cortisol

Version 2 (August 2011)

Contact:
Daniel Powell
Telephone (office): 023 8059 4719
Email: daniel.powell@soton.ac.uk

This guide is intended to remind you of the important aspects of the study, and those functions that are 'built-in' to make things easier for you.

The Handheld you have been given should guide you thoroughly through every step. It should be easy to follow. ***Please keep the Handheld and enough Salivettes for one day with you at all times.***

If you have any problems or questions whatsoever, please contact me by the following means:

Daniel Powell
Telephone: (023) 8059 4719
Email: daniel.powell@soton.ac.uk
(Address of study lab on back page of booklet)

Thank you for participating in this research!

Contents 2

REMINDER: For your own safety, do not attempt to respond to your Handheld while driving.

Day to Day Instructions 3

Morning *(First 45minutes after waking up – Please note you will not be able to postpone a sample within this period)*

It is very important to the study for you to begin as soon as you are awake. If you wake due to disturbed sleep and are going back to sleep do not feel that you need to begin. Go back to sleep and begin when you wake up properly. **You must be awake by 8.30am.**

Each day's sampling/measurement will begin either by:

1. Handheld alarm set the night before by you (*see page 5*)
2. Manually initiating the Handheld upon awakening (immediately switching the Handheld on when you are awake by using the button at the top and then selecting *Start* on the front screen)
3. 8.30am, when the alarm will sound automatically if not already begun

To prevent contamination, please remember to refrain from the following during the first 45mins after awakening:

- Eating
- Drinking (water **is** allowed though)
- Exercise
- Smoking
- Brushing teeth

At each time-point, your handheld will present a **random 3-digit number**. Please transfer this number to the label of the Salivette you are using. Then please complete all questions displayed.

Once these 45 minutes are completed, the Handheld will advise you that you can now eat, drink tea/coffee/juice, etc.

Day to Day Instructions 4

Rest of the Day *(6 further samples + end of day assessment)*

The Handheld will prompt you to take six further samples between 10.30 and 8pm at random intervals spread throughout the day.

Similarly, you will be presented with a **random 3-digit number** each time which you should transfer to the Salivette label. Please answer all questions displayed on the Handheld.

If, when prompted, you have recently eaten or drank alcohol or caffeine, please rinse your mouth with water, empty your mouth, and then wait 20 seconds before taking the sample.

After your 9th sample of the day (3 morning + 6 day), the Handheld will tell you that there will be one final prompt. For this final prompt, you will not be asked to take a saliva sample but will be asked some questions about your experiences of the day overall. This will be the **final alert of the day** (the Handheld will inform you of this too).

The final alert of the day will be **at 9pm**.

Please remember to set the next morning's alarm if you wish to use this function.

Please have the Handheld close-by when you go to bed.

The Handheld 5

The Handheld comes with several functions intended to ease the burden on you, and make things as easy as possible.

➢ ***Quiet mode***

If selected, this will make the next alert quiet (only a quiet short tone will be made).

1. Select '***Help***'.
2. Select '**Quiet mode**'
3. Tick the '***Make Next Alarm Quiet***' box.

NOTE: This will only make ONE (the next) alert quiet. You will have to go back to the Help Menu to make any subsequent alert quiet.

➢ ***Postpone function***

Should you be in a position where you cannot reply to the handheld at that moment (e.g. cooking), it is possible to postpone the alarm for 5, 10, or 15 minutes.

1. When the Handheld alerts you to a sample/measurement, a '***Postpone***' box will appear near the top of the screen. Select.
2. A drop down menu will appear. Select how long you would like to postpone for.

NOTE: Postponement won't be possible in first 45mins of each day.

➢ ***Setting the Morning Alarm***

It is possible to set your morning alarm to a time that suits you (earlier than 8.30am) much like you would with a regular alarm clock.

1. Select '***Help***'.
2. Select '***Morning Alarm***'.
3. Select the Hour and Minute using the up/down arrows.

Please have the Handheld close-by when you go to bed.

Salivettes 6

The Salivettes are plastic vials containing synthetic saliva swabs.

➢ ***To Open***

- Wrap your hand around the Salivette and place your thumb on the top of the lid.
- Slide your thumb to the bottom of the lid and gently flip it upwards. The lid should come off easily.
- Tip the Salivette upside down onto your palm and the swab should fall out.

➢ ***Sampling***

- Place the swab into your mouth.
- Chew gently for approximately 60 seconds until it is saturated with saliva.
- Replace into the Salivette.
- Record the 3-digit number from the Handheld onto the label.
- Place used Salivette into your study bag.

Please store your used Salivettes in a refrigerator at the end of each day.

Closing session appointment time:

Please come to the address below, and remember to bring with you:

- **All Salivettes (4 bags)**
- **Handheld**

The Study Lab is found at:

Room 4035 (Level 4)
Shackleton Building (Building 44)
Highfield Campus
University of Southampton
Southampton, SO17 1BJ

Appendix F – Handheld device homepage screen

Hewlett Packard iPAQ 111 Classic Handheld *(Bracknell, UK*)

Appendix G –Area under the curve (AUC) computations

The formulae presented are based on those given by J. C. Pruessner et al. (2003) to compute the cortisol awakening response (CAR) area under the curve with respect to ground (AUCg) and the area under the curve with respect to increase (AUCi). Each outcome has been divided by 45 so that outcome units correspond to nmol/L/min.

CAR AUCg is considered an outcome representing an estimation of total cortisol secreted during the observation period, in this case 45 minutes. CAR AUCi is considered a measure of the dynamic response to awakening; it measures the estimated cortisol output with respect to the cortisol level at awakening (S1) (Fekedulegn et al., 2007).

Equation

$$AUCg = \frac{\big((S1\ cortisol + S2\ cortisol) * 30\big) + ((S2\ cortisol + S3\ cortisol) * 15)}{45}$$

$$AUCi = AUCg - \frac{S1\ cortisol * (30 + 15)}{45}$$

Where $S1\ cortisol$ represents cortisol level upon awakening, $S2\ cortisol$ represented cortisol level 30 minutes after awakening, and $S3\ cortisol$ represents cortisol level 45 minutes after awakening.

Appendix H – Exploratory factor analysis scree plots

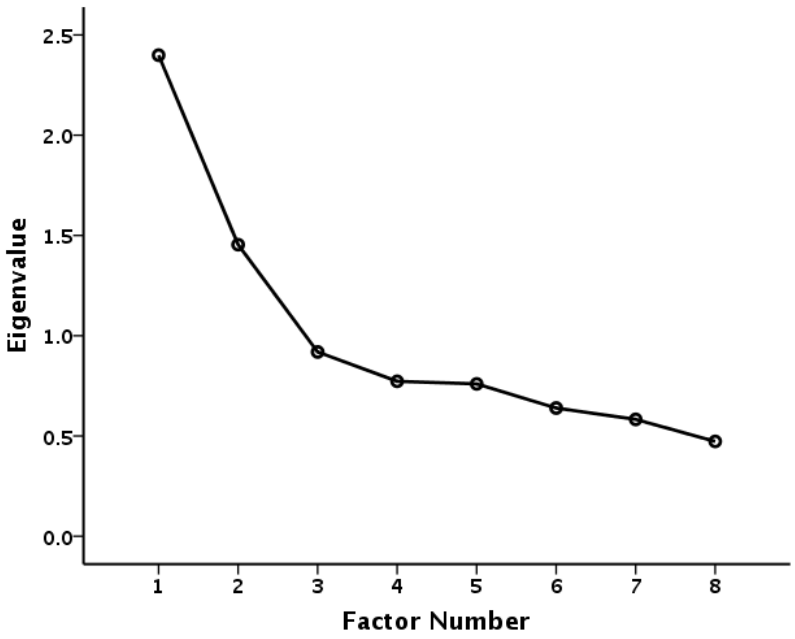


Figure A. Scree plot for momentary stressor items

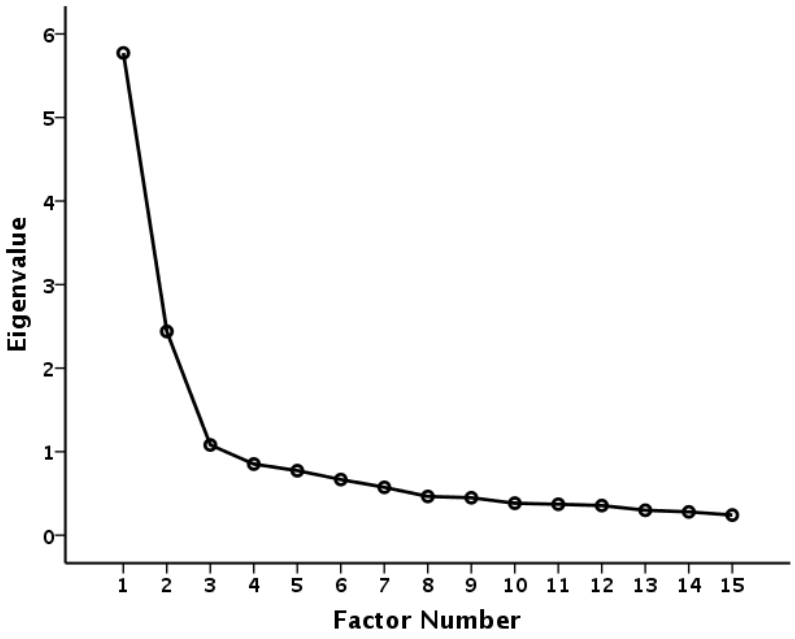


Figure B. Scree plot for momentary mood assessments

Appendix I – Momentary event (A1–A6) completion times

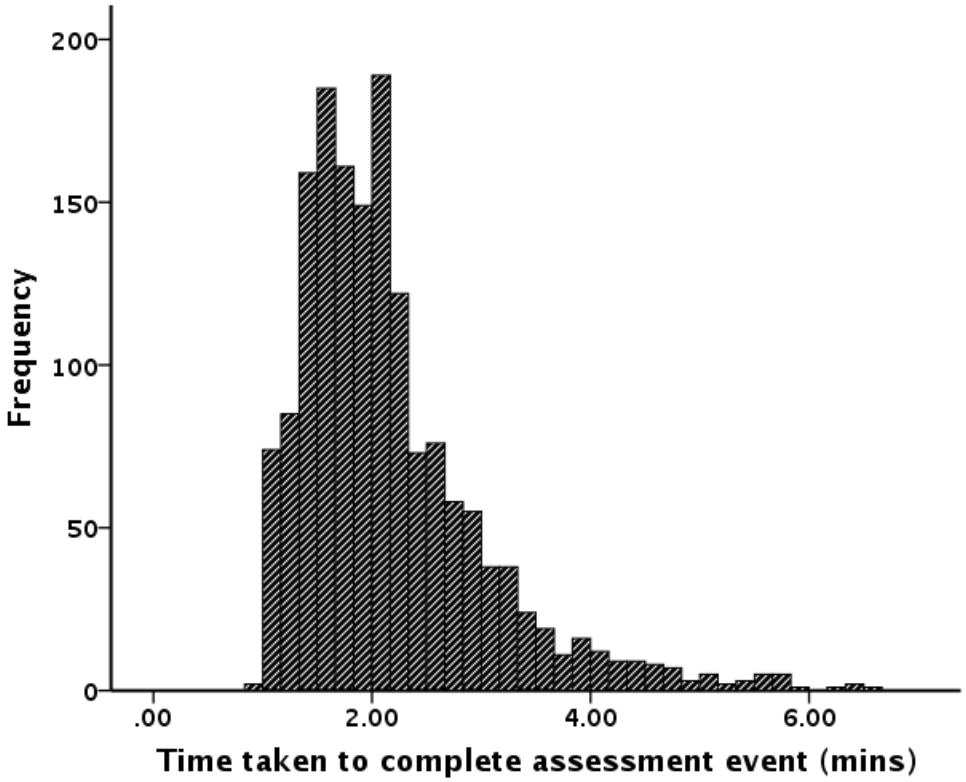


Figure A. Histogram of time taken to complete assessment event

Time taken (mins) in outlying events: 30.00, 28.70, 17.83, 17.25, 16.00, 11.87, 10.00, 9.43, 8.27, 8.06, 8.03, 7.58, 7.53, 7.48, 7.43, 7.38, 6.98, 6.78.

All were excluded.

Appendix J – Multilevel model specifications (momentary fatigue outcome)

Model 1: Unconditional growth model

$$FATIGUE_{adi} = \gamma_{000} + V_{00i} + U_{0di} + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$ and $V_{00i} \sim N(0, \sigma_{00}^2)$

Model 2: Random linear growth model

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + U_{1di}(TIME_{adi}) + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$ and

$\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & \sigma_{0001} \\ \sigma_{1000} & \sigma_{01}^2 \end{bmatrix}\right)$

Model 3: Fixed quadratic random linear growth model

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + U_{1di}(TIME_{adi}) + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$ and

$\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & \sigma_{0001} \\ \sigma_{1000} & \sigma_{01}^2 \end{bmatrix}\right)$

Model 4: Random quadratic growth model

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + V_{00i} + V_{10i}(TIME_{adi}) + V_{20i}(TIME_{adi})^2 + U_{0di} + U_{1di}(TIME_{adi}) + U_{2di}(TIME_{adi})^2 + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \\ U_{2di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$ and

$\begin{bmatrix} V_{00i} \\ V_{10i} \\ V_{20i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & \sigma_{0010} & \sigma_{0020} \\ \sigma_{1000} & \sigma_{10}^2 & \sigma_{1020} \\ \sigma_{2000} & \sigma_{2010} & \sigma_{20}^2 \end{bmatrix} \right)$

Model 5: Random quadratic growth model with quadratic effect fixed at Level-3

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + U_{1di}(TIME_{adi}) + U_{2di}(TIME_{adi})^2 + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \\ U_{2di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$ and

$\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & 0 \\ 0 & \sigma_{10}^2 \end{bmatrix} \right)$

Model 6: Random quadratic covariate growth model with quadratic effect fixed at Level-3

$$FATIGUE_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + \gamma_{300}(EXERTION) + \gamma_{010}(SLEEP\ QUALITY) + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + U_{1di}(TIME_{adi}) + U_{2di}(TIME_{adi})^2 + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $\begin{bmatrix} U_{0di} \\ U_{1di} \\ U_{2di} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$ and

$\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & 0 \\ 0 & \sigma_{10}^2 \end{bmatrix} \right)$

Where $FATIGUE_{adi}$ is the value of MF for person i on day d at assessment a; $TIME_{adi}$the time of day (centred at 1000h) for person i on day d at assessment a. Fixed effects are denoted by γ, with γ_{000} indicating the average intercept, γ_{100} the average effect of *TIME*, γ_{200} the average effect of *TIME*^{2} (quadratic). Random effects at Level-3 (individual) are indicated by V, with V_{00i} denoting the deviation of intercept for person i from the average intercept, V_{10i} denoting the deviation of the effect of person i from the average effect of *TIME*, and V_{20i} denoting the deviation of the effect of person i from the average effect of *TIME*^{2}. Random effects at Level-2 (day) are indicated by U, with U_{0di} denoting the deviation of intercept for day d from the average intercept of person i, U_{1di} denoting the deviation of the effect of day d from the average effect of *TIME* of person i, and U_{2di} denoting the deviation of the effect of day d from the average effect of *TIME*^{2} of person i. R_{adi} indicates residuals at the level of the individual assessments.

Appendix K – Variance/covariance matrices (end-of-day fatigue)

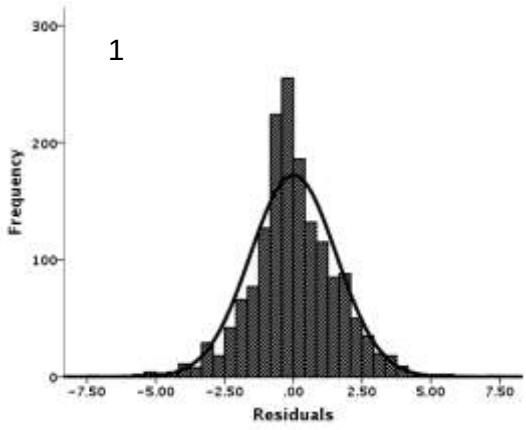
Table End-of-day fatigue variance-covariance structures

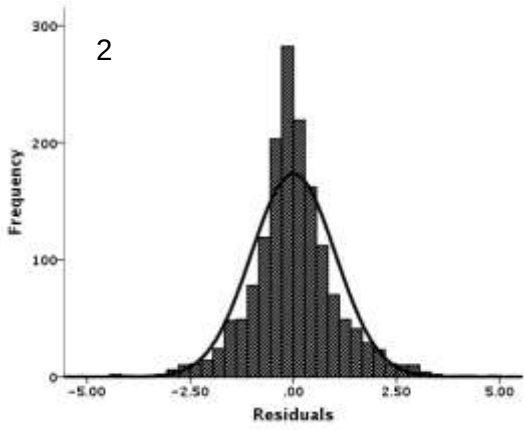
	AIC	BIC
UN	1193.26	1232.43
CS	1194.65	1205.33
CSH	1186.99*	1220.36
AR1	1190.07	1206.76
ARH1	1188.16	1221.53
TP **	1187.41	1205.22*

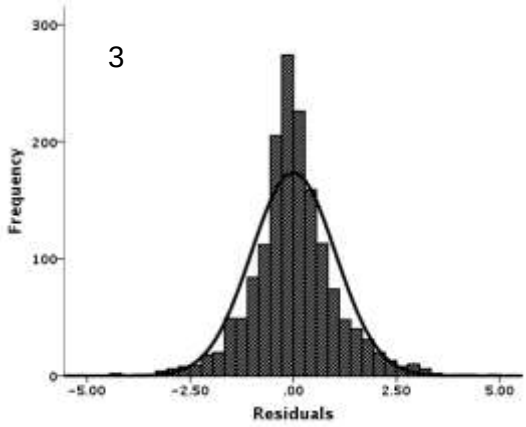
Note: AIC = Akaike information criterion, BIC = Bayesian information criterion, UN = unstructured, CS = compound symmetry, CSH = heterogeneous compound symmetry, AR1 = first order autoregressive, ARH1 = heterogeneous first order autoregressive, TP = toeplitz.
* lowest AIC and BIC score; lowest scores indicate most efficacious model accounting for parsimony.
** chosen variance-covariance structure.

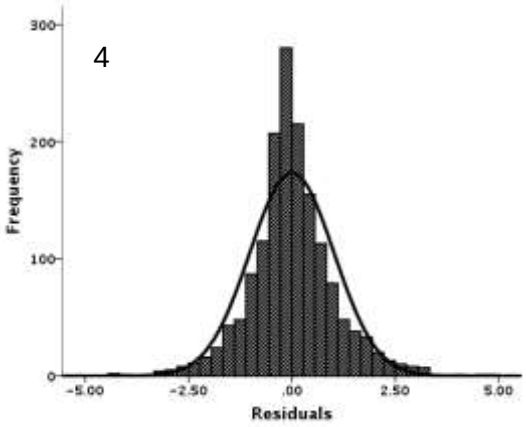
Appendix L – Model residual distributions (Chapter 3)

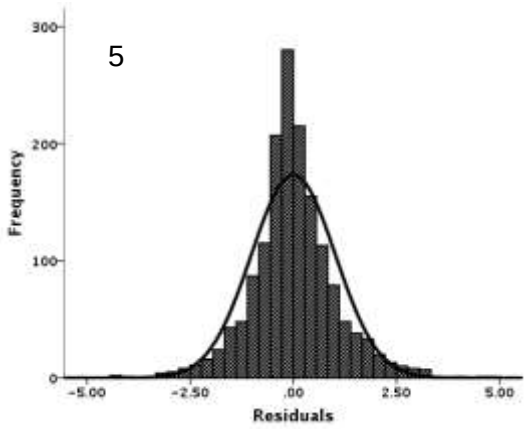
1. Null model (momentary fatigue)
2. Random Linear Fixed Quad covariate model
3. Group comparison (RRMS v Control) diurnal fatigue trends
4. Group comparisons (RRMS-nf as comparator)
5. Group comparisons (Controls as comparator)
6. Peak & End Effects on end-of-day fatigue reports

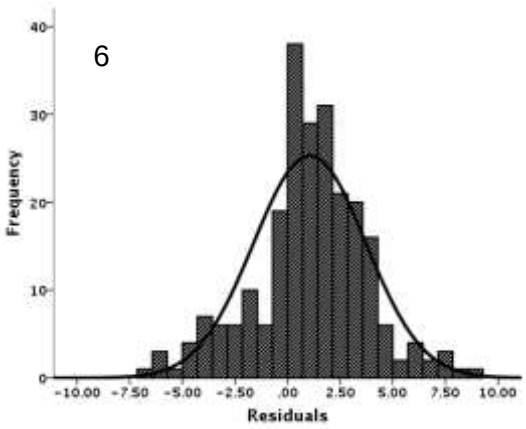












Appendix M – Stressor and mood diurnal patterns

In all graphs, solid line indicates average growth trend for RRMS group; dotted line indicates average for controls group.

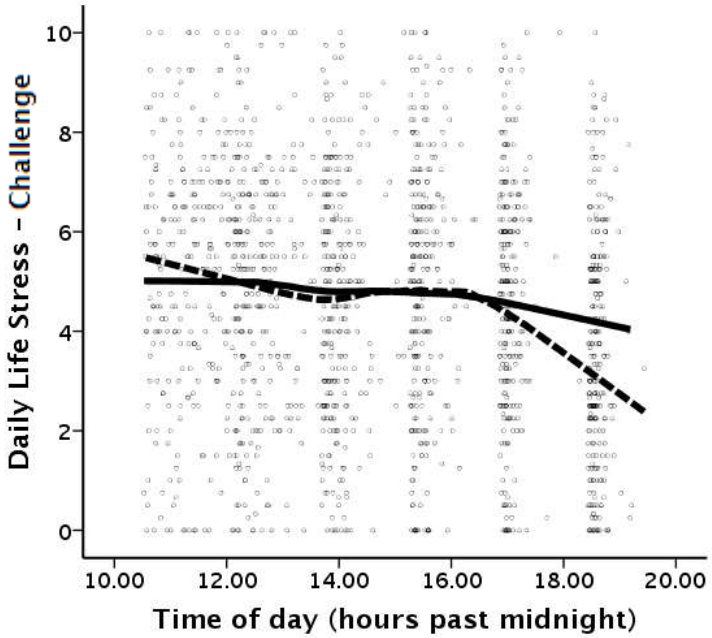


Figure A. Stressor Challenge mean daily trajectories (loess curve)

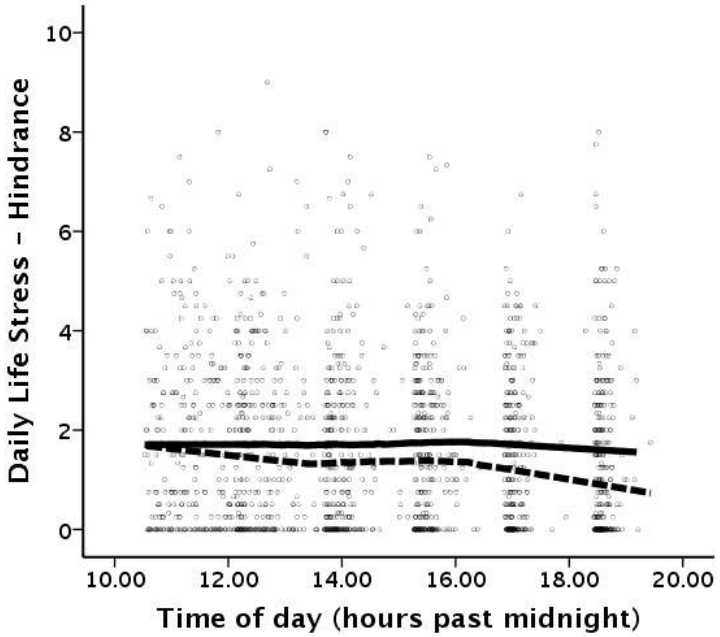


Figure B. Stressor Hindrance mean daily trajectories (loess curve)

Appendix M

In all graphs, solid line indicates average growth trend for RRMS group; dotted line indicates average for controls group.

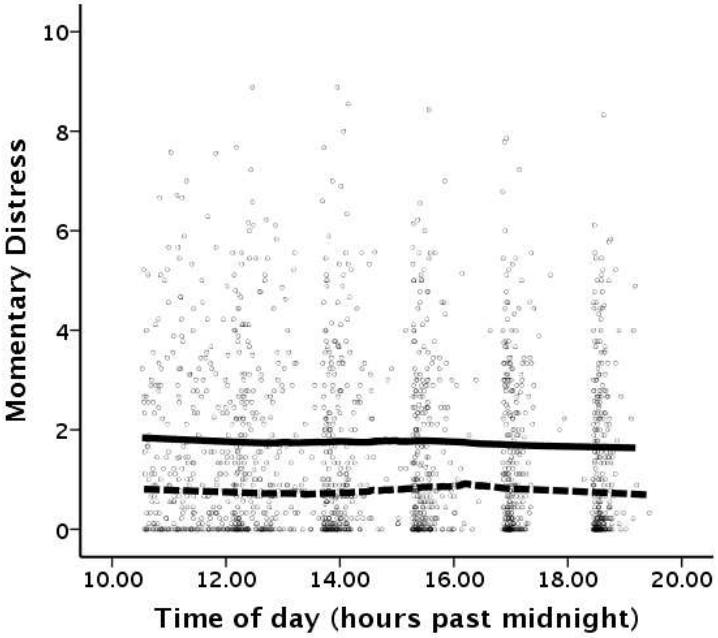


Figure C. Distress mean daily trajectories (loess curve)

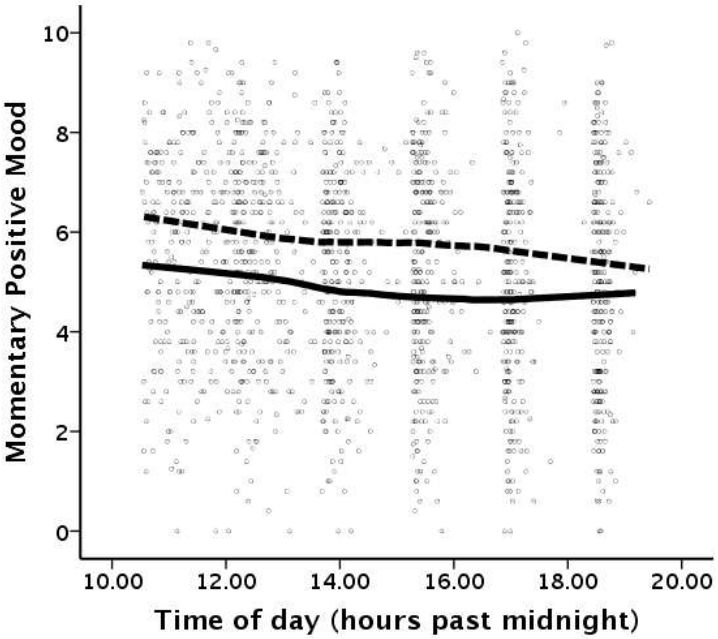


Figure D. Positive Mood mean daily trajectories (loess curve)

Appendix N – Outlier identification (cortisol outcomes)

In the following graphs, circles indicate unique observations. Larger bold diamonds indicate potentially influential outliers on the mean ±3*SD* criterion.

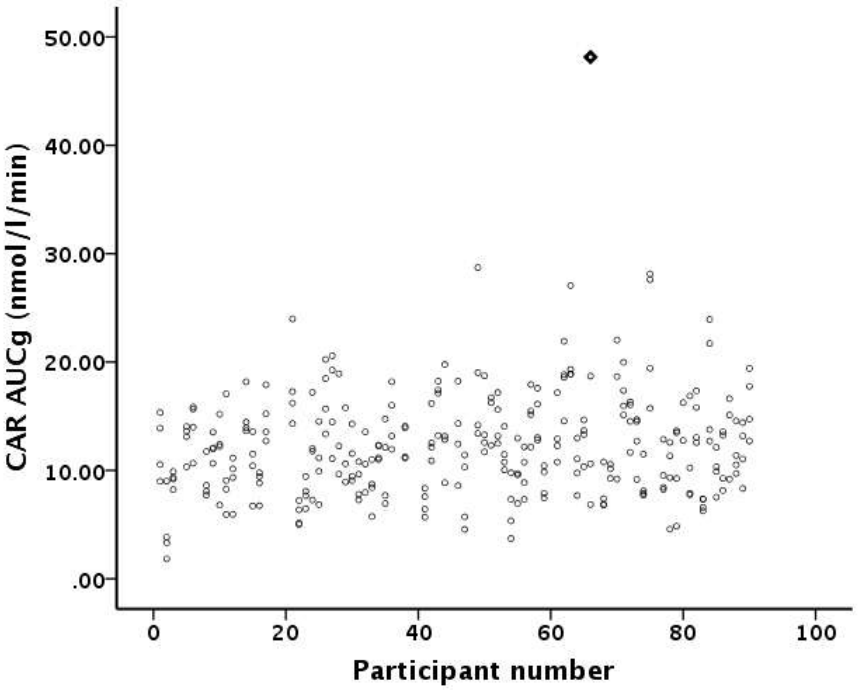


Figure A. Cortisol awakening response area under the curve ground.

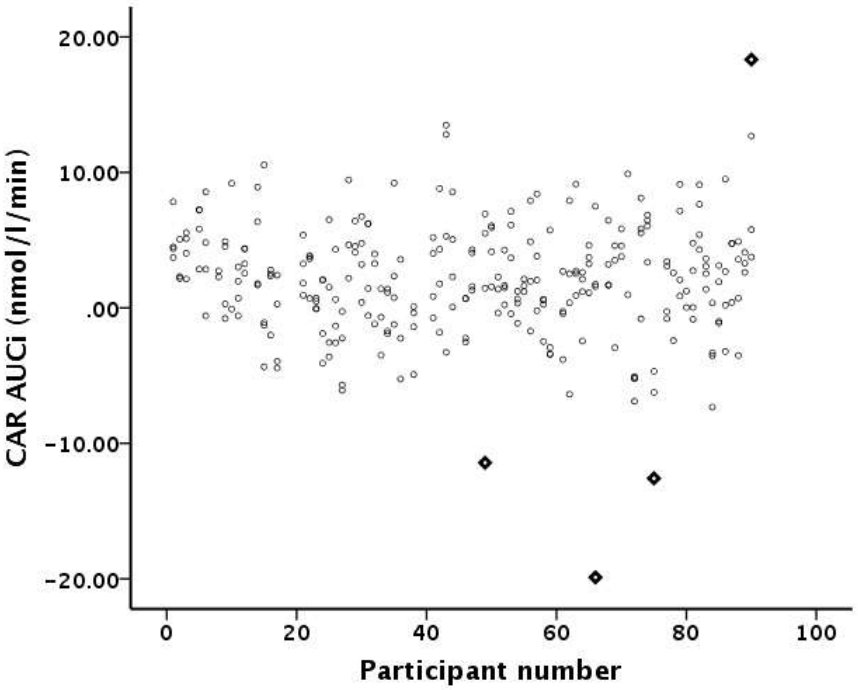


Figure B. Cortisol awakening response area under the curve increase.

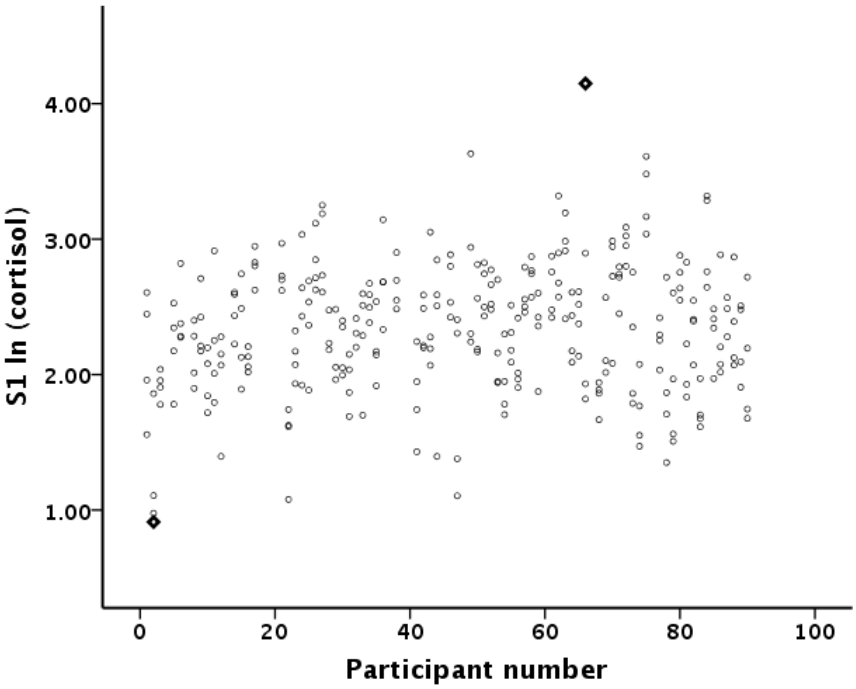


Figure C. S1natuaral-log transformed cortisol.

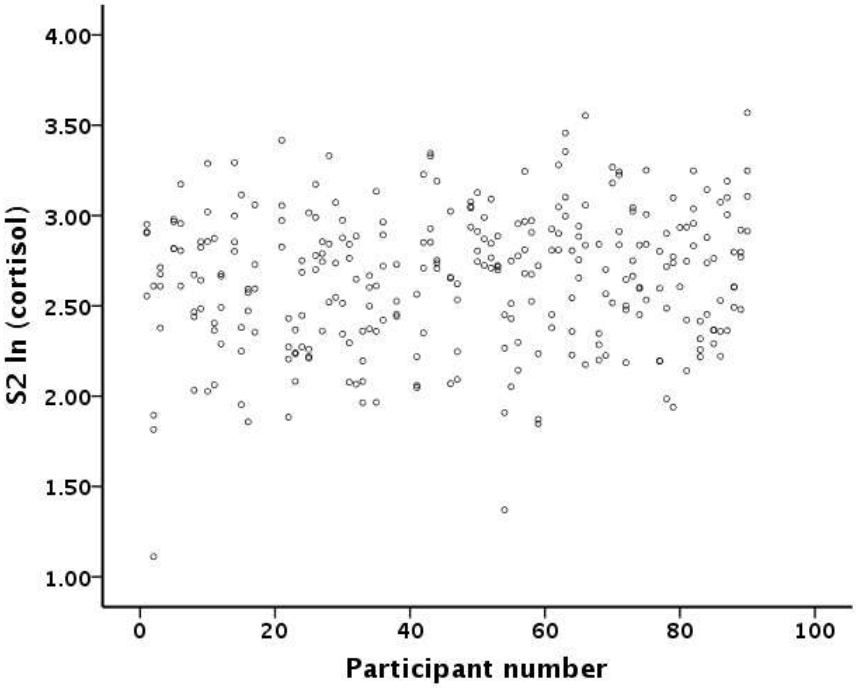


Figure D. S2 natural-log transformed cortisol.

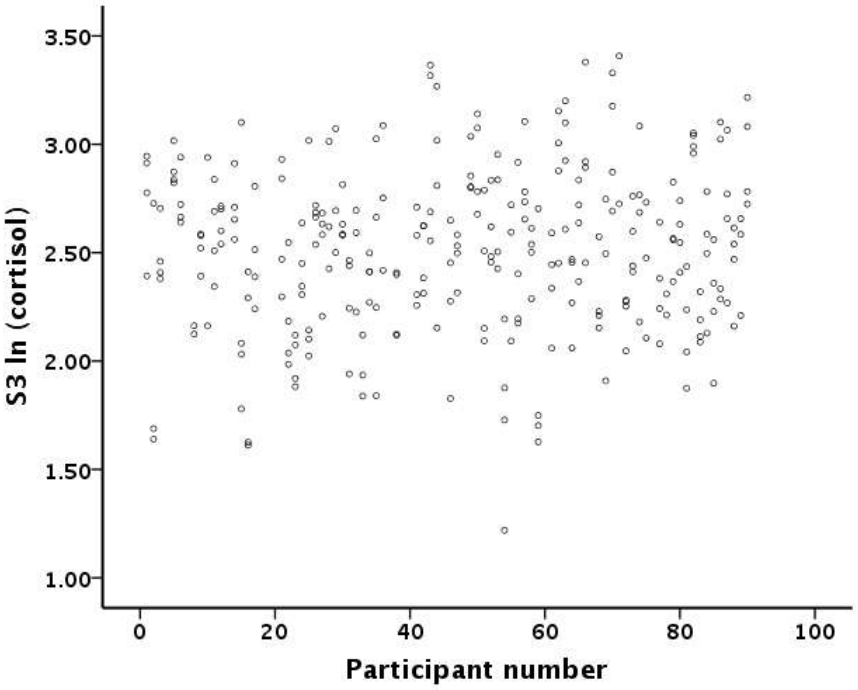


Figure E. S3 natural-log transformed cortisol.

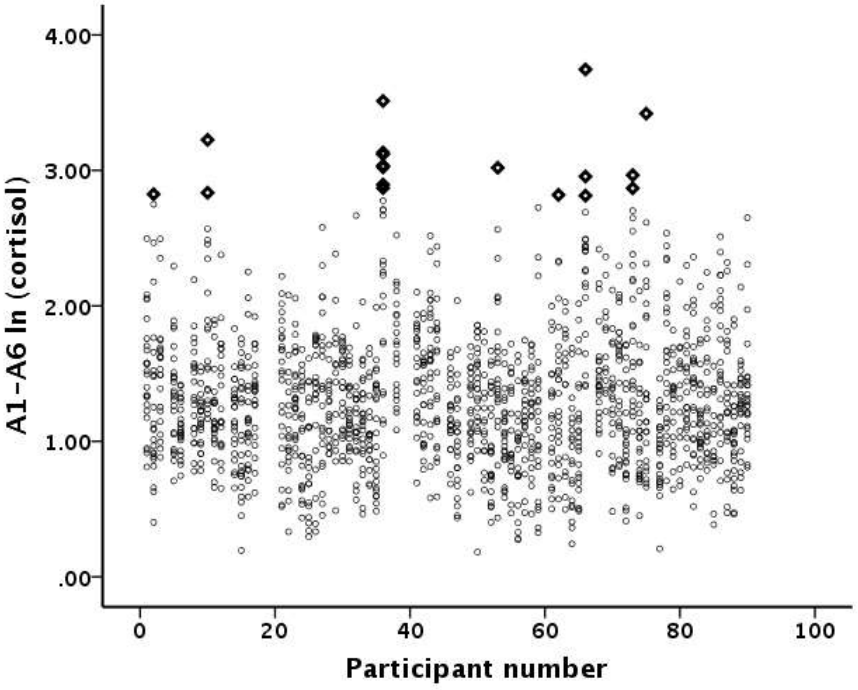


Figure F. Natural-log transformed daytime (1000h – 2000h) cortisol

Appendix O – Model residual distributions (Chapter 4)

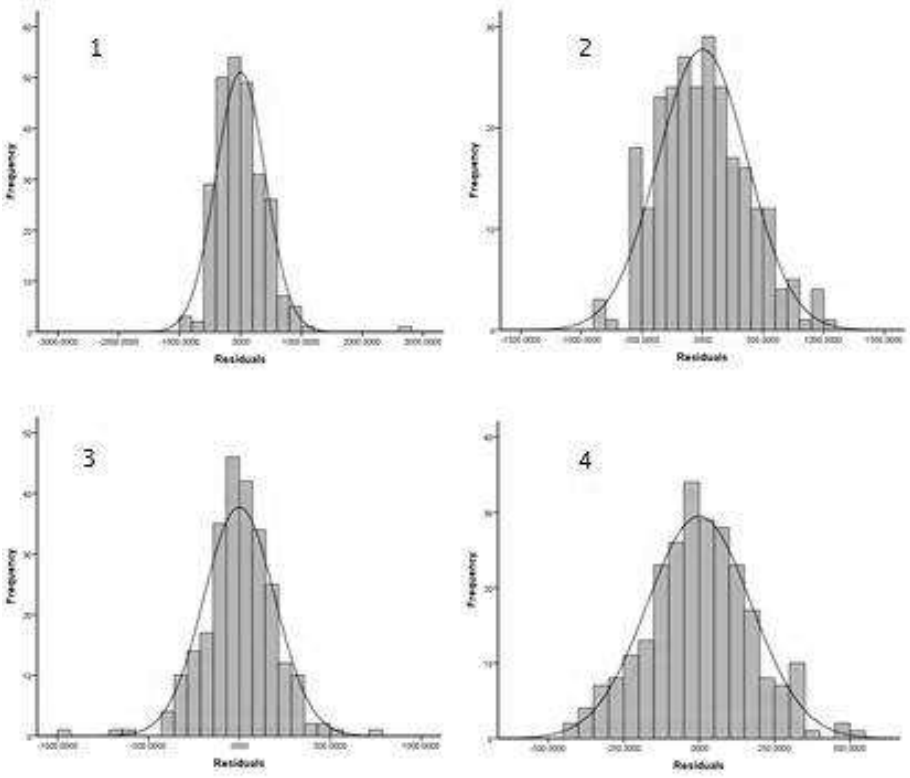


Figure A. Level-1 residuals for group differences with (1) CAR AUCg; (2) CAR AUCg, without outliers; (3) CAR AUCi; (4) CAR AUCi, without outliers.

Appendix O

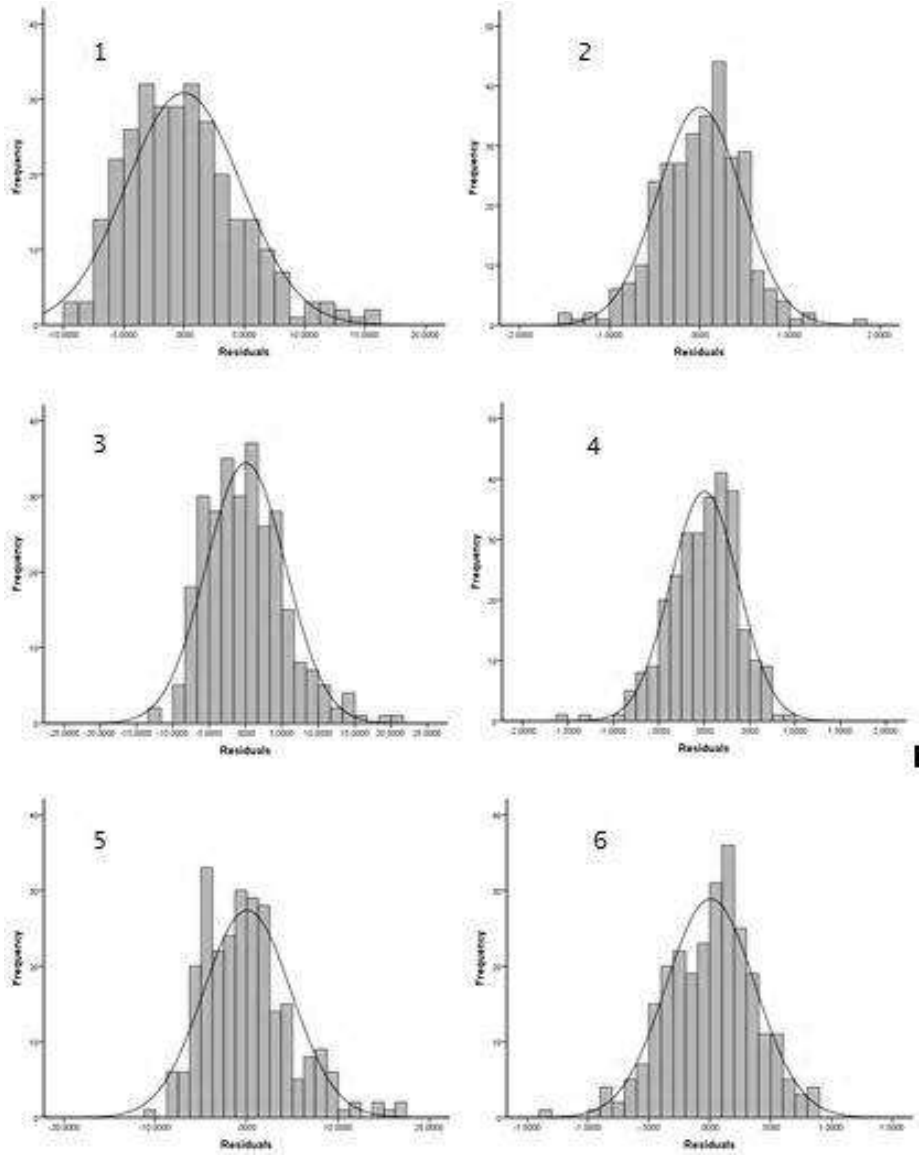


Figure B. Level-1 residuals for group comparisons with (1) S1 cortisol; (2) S1 log-transformed cortisol (3) S2 cortisol; (4) S2 log-transformed cortisol; (5) S3 cortisol; (6) S3 log-transformed cortisol as outcome.

Appendix P – Variance/covariance matrices (cortisol awakening response)

Table A S1 cortisol variance-covariance structures

	AIC	BIC
UN	325.27	365.94
CS **	324.71*	335.81*
CSH	326.15	348.33
AR1	335.23	346.32
ARH1	337.84	360.02
TP	325.29	343.78

Note: AIC = Akaike information criterion, BIC = Bayesian information criterion, UN = unstructured, CS = compound symmetry, CSH = heterogeneous compound symmetry, AR1 = first order autoregressive, ARH1 = heterogeneous first order autoregressive, TP = toeplitz.
* lowest AIC and BIC score; lowest scores indicate most efficacious model accounting for parsimony.

The best-fitting CS variance-covariance matrix is informative as it means the analysis resembles that of a repeated measures ANOVA; there is minimal variability between days and thus no substantial heteroscedasticity (variance is assumed constant at each assessment, and covariance constant between assessments).

Appendix P

Table B CAR AUCi variance-covariance structures

	AIC	BIC
UN	3442.76	3481.84
CS	3442.20	3452.86
CSH	3447.59	3468.91
AR1 **	3432.90*	3443.56*
ARH1	3438.02	3459.34
TP	3435.65	3453.41

Note: AIC = Akaike information criterion, BIC = Bayesian information criterion, UN = unstructured, CS = compound symmetry, CSH = heterogeneous compound symmetry, AR1 = first order autoregressive, ARH1 = heterogeneous first order autoregressive, TP = toeplitz.
* lowest AIC and BIC score; lowest scores indicate most efficacious model accounting for parsimony.

Table C CAR AUCg variance-covariance structures

	AIC	BIC
UN	3732.49	3771.53
CS **	3722.43	3733.08*
CSH	3727.83	3749.12
AR1	3722.58	3733.23
ARH1	3727.73	3749.02
TP	3722.22*	3739.97

Note: AIC = Akaike information criterion, BIC = Bayesian information criterion, UN = unstructured, CS = compound symmetry, CSH = heterogeneous compound symmetry, AR1 = first order autoregressive, ARH1 = heterogeneous first order autoregressive, TP = toeplitz.
* lowest AIC and BIC score; lowest scores indicate most efficacious model accounting for parsimony.

Appendix Q – Multilevel model specification (log-transformed cortisol outcome)

Equation A

Random linear growth model

$$\ln(CORTISOL)_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + R_{adi}$$

Equation B

Random quadratic growth model

$$\ln(CORTISOL)_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + R_{adi}$$

Equation C

Random third polynomial growth model

$$\ln(CORTISOL) = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + \gamma_{300}(TIME_{adi})^3 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + R_{adi}$$

Equation D

Random fourth polynomial growth model

$$\ln(CORTISOL)_{adi} = \gamma_{000} + \gamma_{100}(TIME_{adi}) + \gamma_{200}(TIME_{adi})^2 + \gamma_{300}(TIME_{adi})^3 + \gamma_{400}(TIME_{adi})^4 + V_{00i} + V_{10i}(TIME_{adi}) + U_{0di} + R_{adi}$$

Where $R_{adi} \sim N(0, \sigma_R^2)$ and $U_{0di} \sim N(0, \sigma_0^2)$ and $\begin{bmatrix} V_{00i} \\ V_{10i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00}^2 & \sigma_{0010} \\ \sigma_{1000} & \sigma_{10}^2 \end{bmatrix}\right)$

Where $\ln(CORTISOL)_{adi}$ is the value of natural log transformed cortisol for person i on day d at assessment a; $TIME_{adi}$ the time of day (centred at 1000h) for person i on day d at assessment a. Fixed effects are denoted by γ, with γ_{000} indicating the average intercept, γ_{100} the average effect of *TIME*, γ_{200} the average effect of *TIME*^{2} (quadratic), γ_{300} the average effect of TIME^3, γ_{400} the average effect of TIME^4. Random effects at Level-3 (individual) are indicated by V, with V_{00i} denoting the deviation of intercept for person i from the average intercept, and V_{10i} denoting the deviation of the effect of person i from the average effect of *TIME*.. Random effects at Level-2 (day) are indicated by U, with U_{0di} denoting the deviation of intercept for day d from the average intercept of person i. R_{adi} indicates residuals at the level of the individual assessment.

Appendix R – Loess curves daily cortisol (A1 - A6 events)

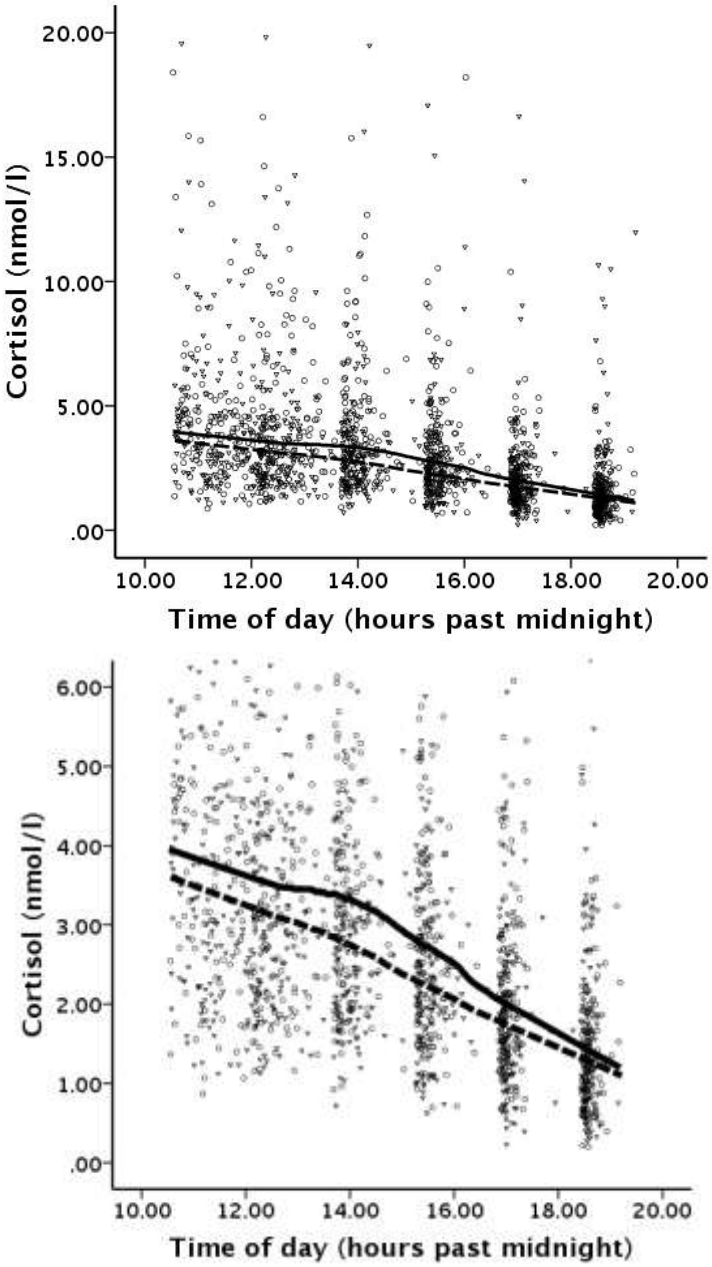


Figure A. Loess curves of cortisol growth curves. Graph in lower panel is identical to upper panel but with a different y-axis scale for observing the slope. Solid line indicates average growth trend for RRMS group; dotted line indicates average for controls group.

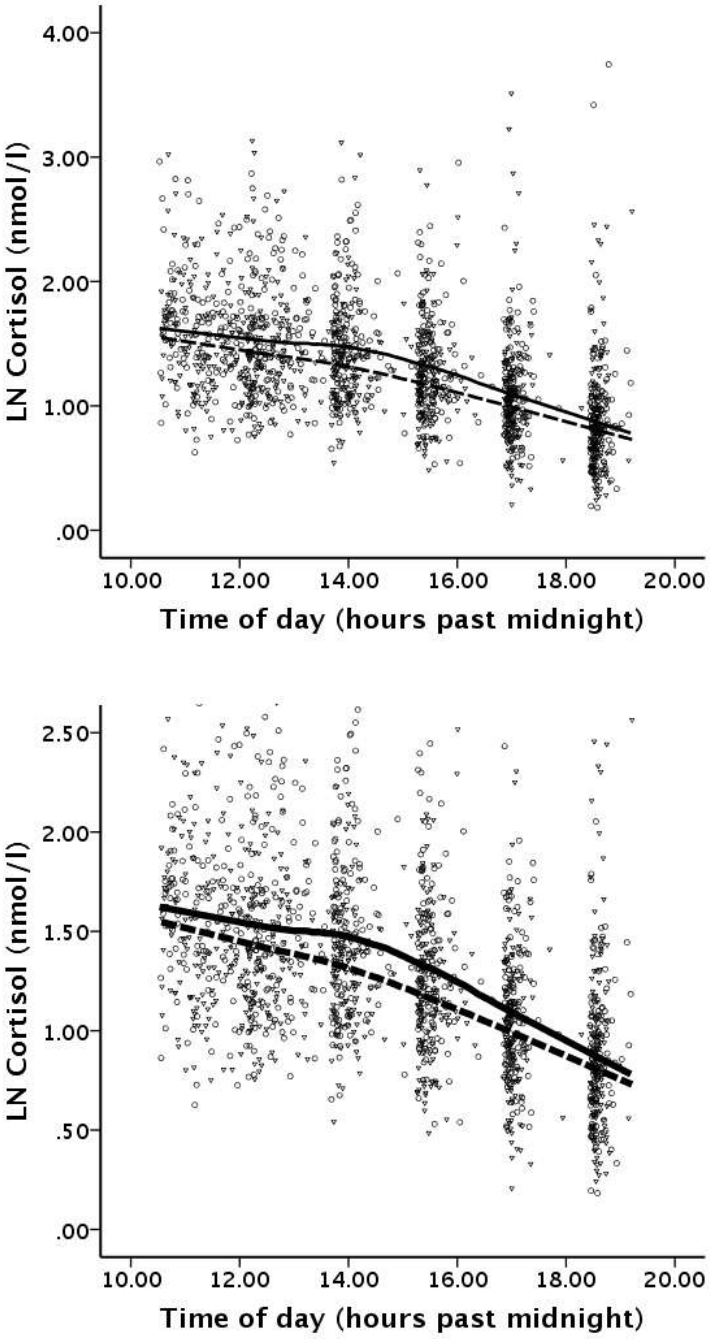


Figure B. Loess curves of natural-log transformed cortisol growth curves. Graph in lower panel is identical to upper panel but with a different y-axis scale for observing the slope. Solid line indicates average growth trend for RRMS group; dotted line indicates average for controls group.

Appendix S – Chronic fatigue syndrome definitions

The Centre for Disease Control and Prevention (CDC) and Oxford case definitions are presented below.

CDC Case Definition (Fukuda et al., 1994)

Patient must satisfy two criteria to be given a chronic fatigue syndrome diagnosis:

- Have self-reported persistent or relapsing fatigue for at least six consecutive months or longer; other medical conditions of which manifestation includes fatigue must be excluded by clinical diagnosis.

- Concurrently have four or more of the following symptoms: post-exertional malaise, impaired memory or concentration, unrefreshing sleep, muscle pain, multi-joint pain without redness or swelling, tender cervical or axillary lymph nodes, sore throat, headache.

The symptoms must have persisted or recurred during six or more consecutive months of illness and must not have predated the fatigue.

Appendix S

Oxford case definition (Sharpe et al., 1991, p. 119)

Chronic fatigue syndrome can be defined:

1. A syndrome characterised by fatigue as the principal symptom.
2. A syndrome of deficit onset that is not life long.
3. The fatigue is severe, disabling, and affects physical and mental functioning.
4. The symptom of fatigue should have been present for a minimum of six months during which it has been present for more than 50% of the time.
5. Other symptoms may be present, particularly myalgia, mood, and sleep disturbance.
6. Certain patients should be excluded from the definition. They include:
   (i) Patients with established medical conditions known to produce chronic fatigue (e.g., severe anaemia). Such patients should be excluded whether the medical condition is diagnosed at presentation or only subsequently. All patients should have a history and physical examination by a competent physician.
   (ii) Patients with a current diagnosis of schizophrenia, manic depressive illness, substance abuse, eating disorder or proven organic brain disease. Other psychiatric disorders (including depressive illness, anxiety disorders, and hyperventilation syndrome) are not necessarily reasons for exclusion.

Appendix T – Systematic review methodological quality scale for studies incorporating salivary cortisol assessments in everyday life

(1) Is the population defined with inclusion and exclusion criteria? (score in brackets)

- Medication use, disease status, psychiatric morbidity; All 3 stated (2)
- Medication use, disease status, psychiatric morbidity; 2 stated (1)
- None or one stated, or not clearly stated (0)

(2) Are the methods for salivary cortisol assessment clearly described and appropriate?

- Two or more assessment days, repeated assessments within days with assessment times reported, saliva sampling method, storage conditions, type of assay performed; All 5 stated (2)
- Repeated assessments within days with assessment times reported, saliva sampling method, storage conditions, type of assay performed; 3-4 stated (1)
- Less than 3 stated or not appropriate (0)

(3) Is adherence to the sampling protocol controlled?

- Electronic monitoring or prompting, with deviations from protocol observed and controlled, including the objectively observed time of awakening (3)
- As above, but without objectively observing the time of awakening (2)
- Electronic prompting, but deviations from protocol not observed and controlled; OR Self-reported sampling times, with deviations observed and controlled (1)
- No appropriate controls, or not stated (0)

(4) If there are early-morning (before 1000h) salivary measures within the design, were all requested at a time relative to the actual waking time (i.e., upon awakening, or awakening plus 30 minutes, etc.) OR more than 60 minutes after awakening?

- Yes, or no pre-1000h measure included (2)
- No, but awakening time assessed and statistically controlled (1)
- No, or not clearly reported (0)

(5) Were missing cortisol assessments dealt with appropriately in the analyses?

- No missing data; OR principled [a] missing data technique used when estimating parameters (3)
- Parameters based on non-complete but adequate data to provide reliable estimates [b] (2)
- Ad hoc [a] missing data technique used (1)
- No method of dealing with missingness reported, or inappropriate (0)

(6) Is the outcome cortisol measurement clearly presented (verbally, graphically or both) with appropriate units?

- Central tendencies, and measures of dispersion presented for each fixed time-point and all computed cortisol estimates (e.g., AUC) (2)
- Central tendency and measures of dispersion presented for either fixed time-point or computed cortisol estimates (1)
- Outcome not clearly presented (0)

(7) Does the study provide appropriate control/adjustment for confounding variables in the relevant analysis? [c]

- Age, gender, socio-economic status, menstrual cycle [d], body mass index, smoking, depression, medication [e], physical exercise, eating shortly before sampling saliva, stressor experience, 6-11 stated (in CAR studies, then also consider waking time, brushing teeth during CAR measurement period, drinking anything other than water during CAR measurement period, sampling day (weekend/weekday), 9-15 stated) (2).
- Age, gender, socio-economic status, menstrual cycle [d], body mass index, smoking, depression, medication [e], physical exercise, eating shortly before sampling saliva, stressor experience, 3-5 stated (in CAR studies, + waking time, brushing teeth during CAR measurement period, drinking anything other than water during CAR measurement period, sampling day (weekend/weekday), 6-8 stated) (1).
- Age, gender, socio-economic status, menstrual cycle [d], body mass index, smoking, depression, medication [e], physical exercise, eating shortly before sampling saliva, stressor experience, 0-2 stated (in CAR studies, + waking time, brushing teeth during CAR measurement period, drinking anything other than water during CAR measurement period, sampling day (weekend/weekday), 0-5 stated) (0).

[a] Principled missing data techniques refer to likelihood-based and Bayesian estimation methods, and multiple imputation. Ad hoc missing data techniques refer to case deletion or single imputation methods.

[b] To be considered an adequate level of data, must have >2 completed assessments for daytime cortisol estimations and >1 assessment for the CAR. Where cases were deleted due to insufficient completed assessments, a comparison of characteristics of included and excluded cases should be made. Not meeting these criteria should result in a score of 0.

[c] Only score for analysis relevant to the review. If study includes any of the confounders as exclusion criteria in participant recruitment, consider these controlled. If potential confounders are compared between groups (with or without explicit matching procedure) and no difference found ($p > .05$), consider that these variables have been controlled for if they are omitted from subsequent analyses. Person-level, day-level, or assessment-level control and adjustment is acceptable.

[d] In male-only studies, menstrual cycle redundant and 6-10 (9-15) required for a score of 2. Requirements for 1 or 0 points unchanged.

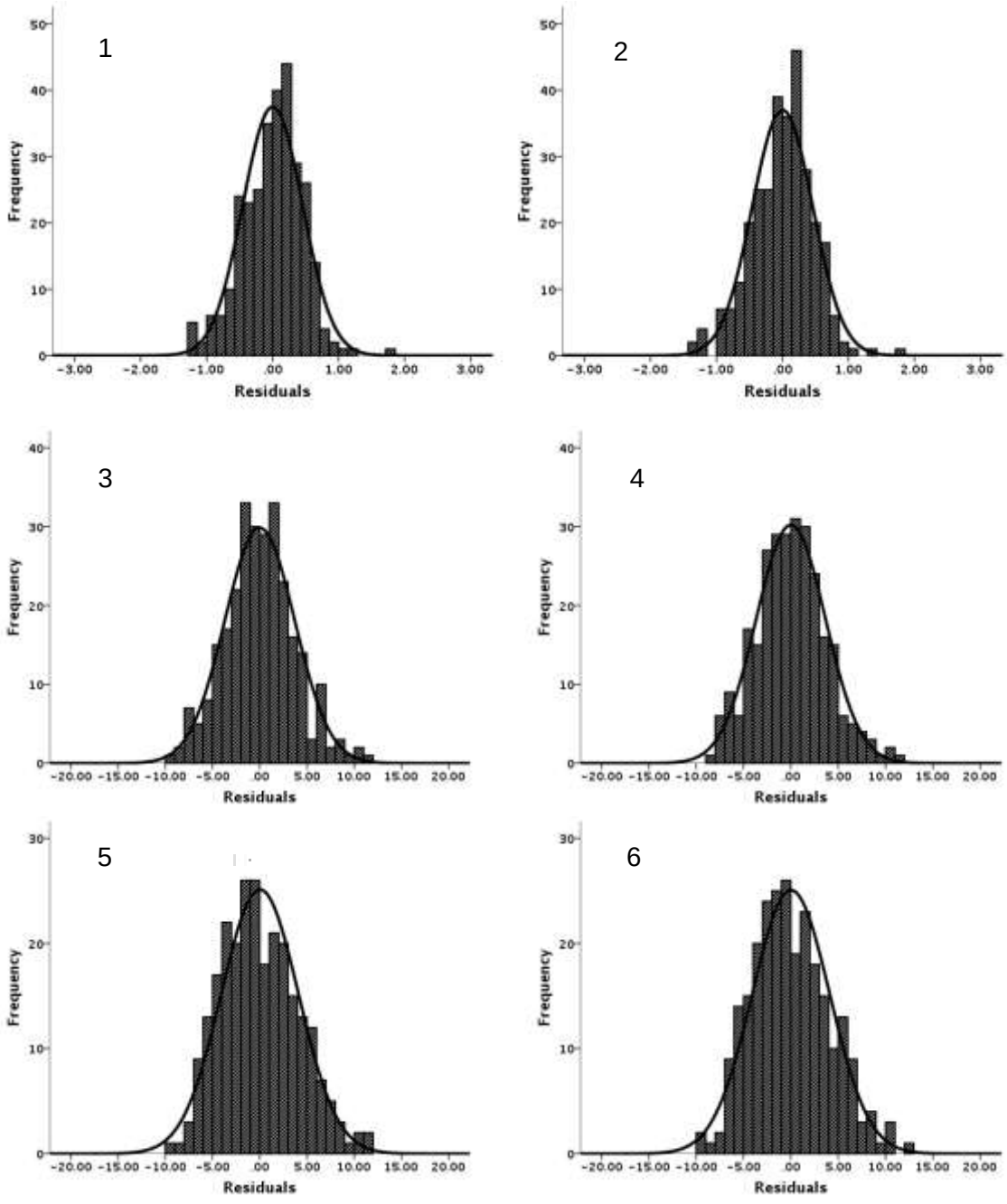
[e] Medication includes hormone replacement therapy, contraceptives, steroids, psychotropic drugs, etc. Accept if study controls for one or all of these.

Appendix U – Systematic review quality scale scores

Study	Quality Scale Item Scores							Total Score
	1	2	3	4	5	6	7	
Chronic fatigue syndrome								
Gaab et al. (2002)	2	2	0	0	0	2	2	8
Jerjes et al. (2005)	2	1	0	0	0	2	1	6
Nater et al. (2008)	2	1	1	2	1	2	2	11
Papadopoulos et al. (2009)	2	0	0	1	0	1	2	6
Rahman et al. (2011)	2	1	0	2	2	2	2	11
Roberts et al. (2004)	2	1	0	2	0	2	2	9
Strickland et al. (1998)	2	1	0	2	0	2	1	8
Torres-Harding et al. (2008)	1	1	0	0	0	0	1	3
A. H. Young et al. (1998)	2	1	0	0	0	2	1	6
Clinical populations								
Barroso et al. (2006)	0	1	0	0	1	2	0	4
Bay and Xie (2009)	1	1	0	0	0	2	2	6
Bower et al. (2005)	2	2	1	0	0	1	2	8
Dekkers et al. (2000)	1	2	1	2	0	1	0	7
Gold et al. (2011)	2	1	0	2	0	2	0	7
McLean et al. (2005)	2	2	3	1	2	2	2	14
Sudhaus et al. (2009)	2	2	0	2	1	2	1	10
Nonclinical populations								
Eek et al. (2012)	0	1	0	2	0	1	1	5
Kumari et al. (2009)	0	1	0	2	2	2	2	9
Lindeberg et al. (2008)	0	1	0	2	0	2	2	10

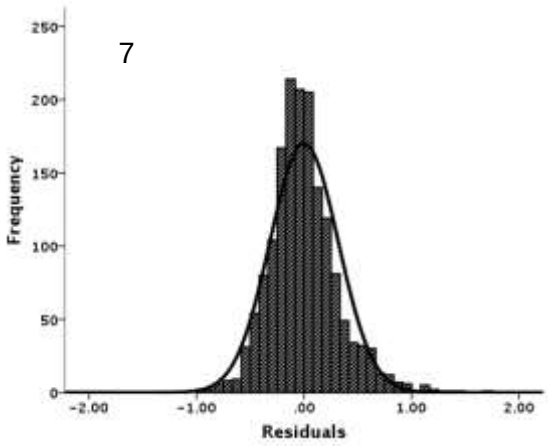
Appendix V – Model residual distributions (Chapter 6)

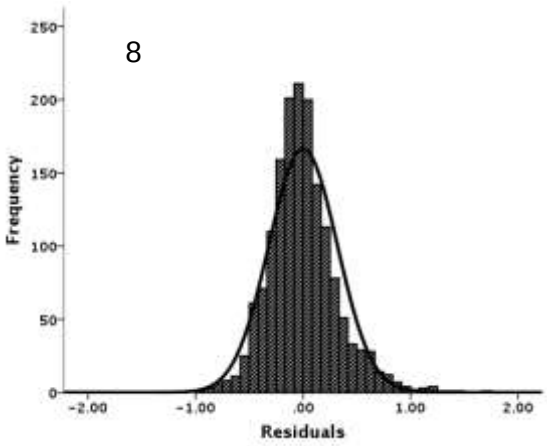
1. S1 cort (Group & CFS predictors)
2. S1 cort (RRMS-f, RRMS-nf, Control)
3. CAR AUCi (Group & CFS predictors)
4. CAR AUCi (RRMS-f, RRMS-nf, Control)
5. CAR AUCg (Group & CFS predictors)
6. CAR AUCg (RRMS-f, RRMS-nf, Control)

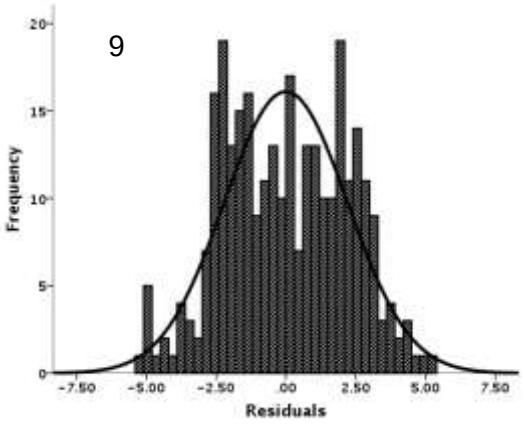


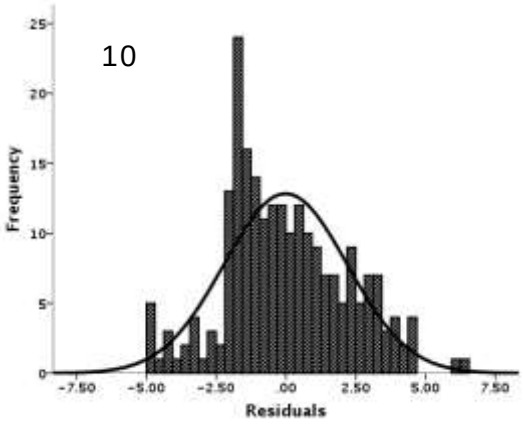
Appendix V

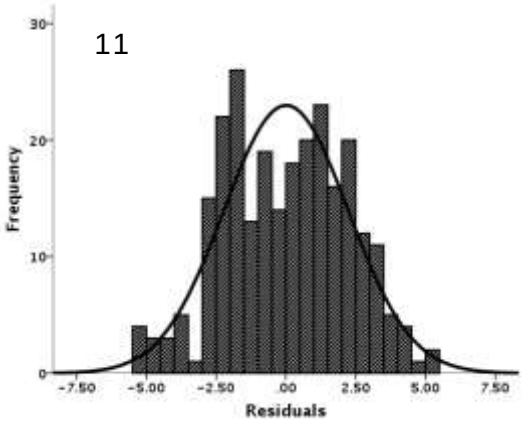
7. Diurnal slopes (fatigue and group)
8. Diurnal slopes (RRMS-f, RRMS-nf, Control)
9. CAR within-subjects (S1 predicting latter-day fatigue)
10. CAR within-subjects (AUCi predicting latter-day fatigue)
11. CAR within-subjects (AUCg predicting latter-day fatigue)











Appendix W – Variance/covariance matrices (daily fatigue severity)

Table Daily mean fatigue covariance structures

	AIC	BIC
UN	1006.47	1045.77
CS**	1003.70	1014.36*
CSH	1001.32*	1022.64
AR1	1019.47	1030.13
ARH1	1017.98	1039.29
TP	1005.55	1023.32

Note: AIC = Akaike information criterion, BIC = Bayesian information criterion, UN = unstructured, CS = compound symmetry, CSH = heterogeneous compound symmetry, AR1 = first order autoregressive, ARH1 = heterogeneous first order autoregressive, TP = toeplitz.
* lowest AIC and BIC score; lowest scores indicate most efficacious model accounting for parsimony.
** chosen covariance structure.

List of References

Aaron, L. A., Turner, J. A., Mancl, L., Brister, H., & Sawchuk, C. N. (2005). Electronic diary assessment of pain-related variables: Is reactivity a problem? *The Journal of Pain, 6*(2), 107-115.

Abercrombie, H. C., Speck, N. S., & Monticelli, R. M. (2006). Endogenous cortisol elevations are related to memory facilitation only in individuals who are emotionally aroused. *Psychoneuroendocrinology, 31*(2), 187-196.

Ackerman, K. D., Martino, M., Heyman, R., & Moyna, N. M. (1996). Immunologic response to acute psychological stress in MS patients and controls. *Journal of Neuroimmunology, 68*(1-2), 85-94.

Ackerman, K. D., Martino, M., Heyman, R., Moyna, N. M., & Rabin, B. S. (1998). Stressor-induced alteration of cytokine production in multiple sclerosis patients and controls. *Psychosomatic Medicine, 60*(4), 484-491.

Adam, E. K., Hawkley, L. C., Kudielka, B. M., & Cacioppo, J. T. (2006). Day-to-day dynamics of experience cortisol associations in a population-based sample of older adults. *Proceedings of the National Academy of Sciences of the United States of America, 103*(45), 17058-17063.

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology, 34*(10), 1423-1436.

Åhsberg, E., Garnberale, F., & Kjellberg, A. (1997). Perceived quality of fatigue during different occupational tasks Development of a questionnaire. *International Journal of Industrial Ergonomics, 20*(2), 121-135.

Aikens, J. E., Reinecke, M. A., Pliskin, N. H., Fischer, J. S., Wiebe, J. S., McCracken, L. M.et al. (1999). Assessing depressive symptoms in multiple sclerosis: Is it necessary to omit items from the original Beck Depression Inventory? *Journal of Behavioral Medicine, 22*(2), 127-142.

Al-Araji, A. H., & Oger, J. (2005). Reappraisal of Lhermitte's sign in multiple sclerosis. *Multiple Sclerosis, 11*(4), 398-402.

Alonso, A., & Hernán, M. A. (2008). Temporal trends in the incidence of multiple sclerosis: a systematic review. *Neurology, 71*(2), 129-135.

References

Altemus, M., Redwine, L., Leong, Y.-M., Yoshikawa, T., Yehuda, R., Detera-Wadleigh, S.et al. (1997). Reduced sensitivity to glucocorticoid feedback and reduced glucocorticoid receptor mrna expression in the luteal phase of the menstrual cycle. *Neuropsychopharmacology, 17*(2), 100-109.

Amato, M. P., Ponziani, G., Rossi, F., Liedl, C. L., Stefanile, C., & Rossi, L. (2001). Quality of life in multiple sclerosis: the impact of depression, fatigue and disability. *Multiple Sclerosis, 7*(5), 340-344.

Amato, M. P., Zipoli, V., Goretti, B., Portaccio, E., De Caro, M. F., Ricchiuti, L.et al. (2006). Benign multiple sclerosis: cognitive, psychological and social aspects in a clinical cohort. *Journal of Neurology, 253*(8), 1054-1059.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th Edition ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR®*: American Psychiatric Pub.

Arnett, P. A., Barwick, F. H., & Beeney, J. E. (2008). Depression in multiple sclerosis: review and theoretical proposal. *Journal of the International Neuropsychological Society, 14*(5), 691-724.

Artemiadis, A. K., Anagnostouli, M. C., & Alexopoulos, E. C. (2011). Stress as a risk factor for multiple sclerosis onset or relapse: A systematic review. *Neuroepidemiology, 36*(2), 109-120.

Artemiadis, A. K., Vervainioti, A. A., Alexopoulos, E. C., Rombos, A., Anagnostouli, M. C., & Darviri, C. (2012). Stress management and multiple sclerosis: A randomized controlled trial. *Archives of Clinical Neuropsychology, 27*(4), 406-416.

Ascherio, A., & Munger, K. L. (2007). Environmental risk factors for multiple sclerosis. Part I: The role of infection. *Annals of Neurology, 61*(4), 288-299.

Ascherio, A., & Munger, K. L. (2010). Epstein–Barr virus infection and multiple sclerosis: A review. *Journal of Neuroimmune Pharmacology, 5*(3), 271-277.

Bach, J.-F. (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. *The New England Journal Of Medicine, 347*(12), 911-920.

Baker, D., Gerritsen, W., Rundle, J., & Amor, S. (2011). Critical appraisal of animal models of multiple sclerosis. *Multiple Sclerosis Journal, 17*(6), 647-657.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology, 22*(3), 309-328.

Bakshi, R. (2003). Fatigue associated with multiple sclerosis: diagnosis, impact and management. *Multiple Sclerosis, 9*(3), 219-227.

Bakshi, R., Miletich, R. S., Henschel, K., Shaikh, Z. A., Janardhan, V., Wasay, M.et al. (1999). Fatigue in multiple sclerosis: cross-sectional correlation with brain MRI findings in 71 patients. *Neurology, 53*(5), 1151.

Bakshi, R., Shaikh, Z. A., Miletich, R. S., Czarnecki, D., Dmochowski, J., Henschel, K.et al. (2000). Fatigue in multiple sclerosis and its relationship to depression and neurologic disability. *Multiple Sclerosis, 6*(3), 181-185.

Balodis, I. M., Wynne-Edwards, K. E., & Olmstead, M. C. (2010). The other side of the curve: Examining the relationship between pre-stressor physiological responses and stress reactivity. *Psychoneuroendocrinology, 35*(9), 1363-1373.

Bamer, A. M., Johnson, K. L., Amtmann, D., & Kraft, G. H. (2008). Prevalence of sleep problems in individuals with multiple sclerosis. *Multiple Sclerosis, 14*(8), 1127-1130.

Bansal, A. S., Bradley, A. S., Bishop, K. N., Kiani-Alikhan, S., & Ford, B. (2012). Chronic fatigue syndrome, the immune system and viral infection. *Brain, Behavior, and Immunity, 26*(1), 24-31.

Baranzini, S. E., Mudge, J., van Velkinburgh, J. C., Khankhanian, P., Khrebtukova, I., Miller, N. A.et al. (2010). Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature, 464*(7293), 1351-1356.

Barroso, J., Burrage, J., Carlson, J., & Carlson, B. W. (2006). Salivary cortisol values in HIV-positive people. *Journal of the Association of Nurses in AIDS Care, 17*(3), 29-36.

Barroso, J., & Lynn, M. R. (2002). Psychometric properties of the HIV-Related Fatigue Scale. *Journal of the Association of Nurses in AIDS Care, 13*(1), 66-75.

References

Bay, E., & Xie, Y. (2009). Psychological and biological correlates of fatigue after mild-to-moderate traumatic brain injury. *Western Journal of Nursing Research, 31*(6), 731-747.

Beatty, W. W., Paul, R. H., Wilbanks, S. L., Hames, K. A., Blanco, C. R., & Goodkin, D. E. (1995). Identifying multiple sclerosis patients with mild or global cognitive impairment using the Screening Examination for Cognitive Impairment (SEFCI). *Neurology, 45*(4), 718-723.

Beck, A. T., Steer, A. J., & Brown, G. K. (1996). *Beck Depression Inventory: Second Edition Manual*. San Antonio, TX: The Psychological Corporation.

Belza, B. L. (1995). Comparison of self-reported fatigue in rheumatoid arthritis and controls. *The Journal of Rheumatology, 22*(4), 639-643.

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. *Cognition & Emotion, 23*(5), 1021-1040.

Beresford, T. P., Arciniegas, D. B., Alfers, J., Clapp, L., Martin, B., Beresford, H. F.et al. (2006). Hypercortisolism in alcohol dependence and its relation to hippocampal volume loss. *Journal of Studies on Alcohol and Drugs, 67*(6), 861-867.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research, 52*(2), 69-77.

Black, A. C., Harel, O., & Matthews, G. (2011). Techniques for analyzing intensive longitudinal data with missing values. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. New York, NY: The Guilford Press.

Blaney, B. E., & Lowe-Strong, A. (2009). The impact of fatigue on communication in multiple sclerosis. The insider's perspective. *Disability and Rehabilitation, 31*(3), 170-180.

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99*(2), 229-246.

Bobholz, J. A., & Rao, S. M. (2003). Cognitive dysfunction in multiple sclerosis: a review of recent developments. *Current Opinion in Neurology, 16*(3), 283-288.

Bol, Y., Duits, A. A., Hupperts, R. M. M., Vlaeyen, J. W. S., & Verhey, F. R. J. (2009). The psychology of fatigue in patients with multiple sclerosis: A review. *Journal of Psychosomatic Research, 66*(1), 3-11.

Bol, Y., Smolders, J., Duits, A., Lange, I. M. J., Romberg-Camps, M., & Hupperts, R. (2012). Fatigue and heat sensitivity in patients with multiple sclerosis. *Acta Neurologica Scandinavica, 126*(6), 384-389.

Bolton, J. E., Humphreys, B. K., & van Hedel, H. J. A. (2010). Validity of weekly recall ratings of average pain intensity in neck pain patients. *Journal of Manipulative and Physiological Therapeutics, 33*(8), 612-617.

Born, J., Hansen, K., Marshall, L., Molle, M., & Fehm, H. L. (1999). Timing the end of nocturnal sleep. *Nature, 397*(6714), 29-30.

Bowen, J., Gibbons, L., Gianas, A., & Kraft, G. H. (2001). Self-administered Expanded Disability Status Scale with functional system scores correlates well with a physician-administered test. *Multiple Sclerosis, 7*(3), 201-206.

Bower, J. E., Ganz, P. A., Dickerson, S. S., Petersen, L., Aziz, N., & Fahey, J. L. (2005). Diurnal cortisol rhythm and fatigue in breast cancer survivors. *Psychoneuroendocrinology, 30*(1), 92-100.

Breitbart, W., McDonald, M. V., Rosenfeld, B., Monkman, N. D., & Passik, S. (1998). Fatigue in ambulatory AIDS patients. *Journal of Pain and Symptom Management, 15*(3), 159-167.

Broderick, J. E., Arnold, D., Kudielka, B. M., & Kirschbaum, C. (2004). Salivary cortisol sampling compliance: comparison of patients and healthy volunteers. *Psychoneuroendocrinology, 29*(5), 636-650.

Broderick, J. E., Schwartz, J. E., Schneider, S., & Stone, A. A. (2009). Can end-of-day reports replace momentary assessment of pain and fatigue? *The Journal of Pain, 10*(3), 274-281.

Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A. A. (2008). The accuracy of pain and fatigue items across different reporting periods. *Pain, 139*(1), 146-157.

Brønnum-Hansen, H., Koch-Henriksen, N., & Stenager, E. (2004). Trends in survival and cause of death in Danish patients with multiple sclerosis. *Brain, 127*(4), 844-850.

Brown, G. W., & Harris, T. (1989). *Life events and illness*. New York, NY: Guilford Press.

Brown, R. F., Valpiani, E. M., Tennant, C. C., Dunn, S. M., Sharrock, M., Hodgkinson, S.et al. (2009). Longitudinal assessment of anxiety, depression, and fatigue in people with multiple sclerosis. *Psychology and Psychotherapy: Theory, Research and Practice, 82*(1), 41-56.

References

Buchanan, T. W., Kern, S., Allen, J. S., Tranel, D., & Kirschbaum, C. (2004). Circadian regulation of cortisol after hippocampal damage in humans. *Biological Psychiatry, 56*(9), 651-656.

Burton, J. M., O'Connor, P. W., Hohol, M., & Beyene, J. (2012). Oral versus intravenous steroids for treatment of relapses in multiple sclerosis. *Cochrane Database of Systematic Reviews*(12), CD006921.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research, 28*(2), 193-213.

Carr, B. R., Parker, C. R., Jr., Madden, J. D., MacDonald, P. C., & Porter, J. C. (1981). Maternal plasma adrenocorticotropin and cortisol relationships throughout human pregnancy. *American Journal of Obstetrics and Gynecology, 139*(4), 416-422.

Cella, M., Chalder, T., & White, P. D. (2011). Does the heterogeneity of chronic fatigue syndrome moderate the response to cognitive behaviour therapy? An exploratory study. *Psychotherapy and Psychosomatics, 80*(6), 353-358.

Cella, M., White, P. D., Sharpe, M., & Chalder, T. (2013). Cognitions, behaviours and co-morbid psychiatric diagnoses in patients with chronic fatigue syndrome. *Psychological Medicine, 43*(02), 375-380.

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D.et al. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research, 37*(2), 147-153.

Charcot, J. M. (1887). *Lectures on the diseases of the nervous system delivered at La Salpêtrière*. London, UK: New Sydenham Society.

Chaudhuri, A., & Behan, P. O. (2000). Fatigue and basal ganglia. *Journal of the Neurological Sciences, 179*(1–2), 34-42.

Chaudhuri, A., & Behan, P. O. (2004). Fatigue in neurological disorders. *Lancet, 363*(9413), 978-988.

Chaudhuri, A., & Behan, P. O. (2005). Multiple sclerosis: looking beyond autoimmunity. *Journal of the Royal Society of Medicine, 98*(7), 303-306.

Chida, Y., & Steptoe, A. (2009). Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. *Biological Psychology, 80*(3), 265-278.

Chwastiak, L. A., Gibbons, L. E., Ehde, D. M., Sullivan, M., Bowen, J. D., Bombardier, C. H.et al. (2005). Fatigue and psychiatric illness in a large community sample of persons with multiple sclerosis. *Journal of Psychosomatic Research, 59*(5), 291-298.

Clark, D. M., & Teasdale, J. D. (1985). Constraints on the effects of mood on memory. *Journal of Personality and Social Psychology, 48*(6), 1595-1608.

Claros-Salinas, D., Bratzke, D., Greitemann, G., Nickisch, N., Ochs, L., & Schröter, H. (2010). Fatigue-related diurnal variations of cognitive performance in multiple sclerosis and stroke patients. *Journal of the Neurological Sciences, 295*(1–2), 75-81.

Cleare, A. J. (2003). The neuroendocrinology of chronic fatigue syndrome. *Endocrine Reviews, 24*(2), 236-252.

Cleare, A. J. (2004). The HPA axis and the genesis of chronic fatigue syndrome. *Trends in Endocrinology and Metabolism, 15*(2), 55-59.

Cleare, A. J., Heap, E., Malhi, G. S., Wessely, S., O'Keane, V., & Miell, J. (1999). Low-dose hydrocortisone in chronic fatigue syndrome: a randomised crossover trial. *The Lancet, 353*(9151), 455-458.

Clow, A., Edwards, S., Owen, G., Evans, G., Evans, P., Hucklebridge, F.et al. (2006). Post-awakening cortisol secretion during basic military training. *International Journal of Psychophysiology, 60*(1), 88-94.

Clow, A., Hucklebridge, F., Stalder, T., Evans, P., & Thorn, L. (2010). The cortisol awakening response: More than a measure of HPA axis function. *Neuroscience & Biobehavioral Reviews, 35*(1), 97-103.

Clow, A., Thorn, L., Evans, P., & Hucklebridge, F. (2004). The awakening cortisol response: Methodological issues and significance. *Stress, 7*(1), 29-37.

Codella, M., Rocca, M. A., Colombo, B., Rossi, P., Comi, G., & Filippi, M. (2002). A preliminary study of magnetization transfer and diffusion tensor MRI of multiple sclerosis patients with fatigue. *Journal of Neurology, 249*(5), 535-537.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

References

Cohen, S., & Janicki-Deverts, D. (2012). Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. *Journal of Applied Social Psychology, 42*(6), 1320-1334.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385-396.

Colombo, B., Boneschi, F. M., Rossi, P., Rovaris, M., Maderna, L., Filippi, M.et al. (2000). MRI and motor evoked potential findings in nondisabled multiple sclerosis patients with and without symptoms of fatique. *Journal of Neurology, 247*(7), 506-509.

Comi, G., Filippi, M., & Wolinsky, J. S. (2001). European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging-measured disease activity and burden in patients with relapsing multiple sclerosis. *Annals of Neurology, 49*(3), 290-297.

Comi, G., Leocani, L., Rossi, P., & Colombo, B. (2001). Physiopathology and treatment of fatigue in multiple sclerosis. *Journal of Neurology, 248*(3), 174-179.

Compston, A., & Coles, A. (2008). Multiple sclerosis. *The Lancet, 372*(9648), 1502-1517.

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine, 74*(4), 327-337.

Cook, J. D. (2005). Diagnosis and management of iron-deficiency anaemia. *Best Practice & Research Clinical Haematology, 18*(2), 319-332.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factory analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research, & Evaluation, 10*(7), 1-9.

Cox, A., Pyne, D., Gleson, M., & Callister, R. (2008). Resting plasma and salivary IL-6 concentrations are not correlated in distance runners. *European Journal of Applied Physiology, 103*(4), 477-479.

Crawford, J. R., Henry, J. D., Crombie, C., & Taylor, E. P. (2001). Normative data for the HADS from a large non-clinical sample. *British Journal of Clinical Psychology, 40*(4), 429-434.

Crow, R., & Hammond, S. (1996). A qualitative study to explore the concept of fatigue/tiredness in cancer patients and in healthy individuals. *Supportive Care in Cancer, 4*(2), 82-96.

Cruise, C. E., Broderick, J. E., Porter, L., Kaell, A., & Stone, A. A. (1996). Reactive effects of diary self-assessment in chronic pain patients. *Pain, 67*(2-3), 253-258.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the Experience-Sampling Method. *Journal of Nervous and Mental Disease, 175*(9), 526-536.

Cullen, W., Kearney, Y., & Bury, G. (2002). Prevalence of fatigue in general practice. *Irish Journal of Medical Science, 171*(1), 10-12.

Curran, S. L., Beacham, A. O., & Andrykowski, M. A. (2004). Ecological momentary assessment of fatigue following breast cancer treatment. *Journal of Behavioral Medicine, 27*(5), 425-444.

Dallman, M. F., la Fleur, S. E., Pecoraro, N. C., Gomez, F., Houshyar, h., & Akana, S. F. (2004). Minireview: Glucocorticoids - food intake, abdominal obesity, and wealthy nations in 2004. *Endocrinology, 145*(6), 2633-2638.

Dalton, E. J., & Heinrichs, R. W. (2005). Depression in multiple sclerosis: A quantitative review of the evidence. *Neuropsychology, 19*(2), 152-158.

de Jong, K., Moerbeek, M., & van der Leeden, R. (2009). A priori power analysis in longitudinal three-level multilevel models: An example with therapist effects. *Psychotherapy Research, 20*(3), 273-284.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R.et al. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*(1), 69-102.

Dekkers, J. C., Geenen, R., Godaert, G. L. R., van Doornen, L. J. P., & Bijlsma, J. W. J. (2000). Diurnal courses of cortisol, pain, fatigue, negative mood, and stiffness in patients with recently diagnosed rheumatoid arthritis. *International Journal of Behavioral Medicine, 7*(4), 353-371.

Demitrack, M. A., Dale, J. K., Straus, S. E., Laue, L., Listwak, S. J., Kruesi, M. J. P.et al. (1991). Evidence for impaired activation of the hypothalamic-pituitary-adrenal axis in patients with chronic fatigue syndrome. *The Journal of Clinical Endocrinology & Metabolism, 73*(6), 1224-1234.

References

Dennison, L., Moss-Morris, R., Silber, E., Galea, I., & Chalder, T. (2010). Cognitive and behavioural correlates of different domains of psychological adjustment in early-stage multiple sclerosis. *Journal of Psychosomatic Research, 69*(4), 353-361.

Derogatis, L. R. (1977). *Symptom Checklist R-version Manual I: Scoring, Administration, and Procedures for the SCL-90*. Baltimore, MD: John Hopkins Press.

Deuschle, M., Gotthardt, U., Schweiger, U., Weber, B., Körner, A., Schmider, J.et al. (1997). With aging in humans the activity of the hypothalamus-pituitary-adrenal system increases and its diurnal amplitude flattens. *Life Sciences, 61*(22), 2239-2246.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355-391.

Dluhy, R. G. (1998). Clinical relevance of inhaled corticosteroids and HPA axis suppression. *Journal of Allergy and Clinical Immunology, 101*(4, Supplement), S447-S450.

Dockray, S., Bhattacharyya, M. R., Molloy, G. J., & Steptoe, A. (2008). The cortisol awakening response in relation to objective and subjective measures of waking in the morning. *Psychoneuroendocrinology, 33*(1), 77-82.

Dressendörfer, R. A., Kirschbaum, C., Rohde, W., Stahl, F., & Strasburger, C. J. (1992). Synthesis of a cortisol-biotin conjugate and evaluation as a tracer in an immunoassay for salivary cortisol measurement. *The Journal of Steroid Biochemistry and Molecular Biology, 43*(7), 683-692.

Dworkin, R. H., O'Connor, A. B., Backonja, M., Farrar, J. T., Finnerup, N. B., Jensen, T. S.et al. (2007). Pharmacologic management of neuropathic pain: Evidence-based recommendations. *Pain, 132*(3), 237-251.

Earley, C. J. (2003). Restless legs syndrome. *New England Journal of Medicine, 348*(21), 2103-2109.

Ebner-Priemer, U. W., & Kubiak, T. (2010). The decade of behavior revisited: Future prospects for ambulatory assessment. *European Journal of Psychological Assessment, 26*(3), 151-153.

Edwards, S., Evans, P., Hucklebridge, F., & Clow, A. (2001). Association between time of awakening and diurnal cortisol secretory activity. *Psychoneuroendocrinology, 26*(6), 613-622.

Eek, F., Karlson, B., Garde, A. H., Hansen, Å. M., & Ørbæk, P. (2012). Cortisol, sleep, and recovery – Some gender differences but no straight associations. *Psychoneuroendocrinology, 37*(1), 56-64.

Engert, V., Efanov, S. I., Duchesne, A., Vogel, S., Corbo, V., & Pruessner, J. C. (2013). Differentiating anticipatory from reactive cortisol responses to psychosocial stress. *Psychoneuroendocrinology, 38*(8), 1328-1337.

Evers, A. W. M., Verhoeven, E. W. M., van Middendorp, H., Sweep, F. C. G. J., Kraaimaat, F. W., Donders, A. R. T.et al. (2013). Does stress affect the joints? Daily stressors, stress vulnerability, immune and HPA axis activity, and short-term disease and symptom fluctuations in rheumatoid arthritis. *Annals of the Rheumatic Diseases*.

Fassbender, K., Schmidt, R., Mössner, R., Kischka, U., Kühnen, J., Schwartz, A.et al. (1998). Mood disorders and dysfunction of the hypothalamic-pituitary-adrenal axis in multiple sclerosis: association with cerebral inflammation. *Archives of Neurology, 55*(1), 66-72.

Fekedulegn, D. B., Andrew, M. E., Burchfiel, C. M., Violanti, J. M., Hartley, T. A., Charles, L. E.et al. (2007). Area under the curve and other summary indicators of repeated waking cortisol measurements. *Psychosomatic Medicine, 69*(7), 651-659.

Fernandez-Botran, R., Miller, J. J., Burns, V. E., & Newton, T. L. (2011). Correlations among inflammatory markers in plasma, saliva and oral mucosal transudate in post-menopausal women with past intimate partner violence. *Brain, Behavior, and Immunity, 25*(2), 314-321.

Feys, P., Gijbels, D., Romberg, A., Santoyo, C., Gebara, B., Maertens de Noordhout, B.et al. (2012). Effect of time of day on walking capacity and self-reported fatigue in persons with multiple sclerosis: a multi-center trial. *Multiple Sclerosis Journal, 18*(3), 351-357.

Field, A., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I Disorders*. Washington, DC: American Psychiatric Press.

Fishbain, D. A., Cutler, R. B., Cole, B., Lewis, J., Smets, E., Rosomoff, H. L.et al. (2004). Are patients with chronic low back pain or chronic neck pain fatigued? *Pain Medicine, 5*(2), 187-195.

References

Fisk, J. D., Pontefract, A., Ritvo, P. G., Archibald, C. J., & Murray, T. J. (1994). The impact of fatigue on patients with multiple sclerosis. *The Canadian Journal of Neurological Sciences, 21*(1), 9-14.

Fisk, J. D., Ritvo, P. G., Ross, L., Haase, D. A., Marrie, T. J., & Schlech, W. F. (1994). Measuring the functional impact of fatigue: Initial validation of the Fatigue Impact Scale. *Clinical Infectious Diseases, 18*(Supplement 1), S79-S83.

Flachenecker, P., Bihler, I., Weber, F., Gottschalk, M., Toyka, K. V., & Rieckmann, P. (2004). Cytokine mRNA expression in patients with multiple sclerosis and fatigue. *Multiple Sclerosis, 10*(2), 165-169.

Flachenecker, P., Kümpfel, T., Kallmann, B., Gottschalk, M., Grauer, O., Rieckmann, P.et al. (2002). Fatigue in multiple sclerosis: a comparison of different rating scales and correlation to clinical parameters. *Multiple Sclerosis 8*(6), 523-526.

Flachenecker, P., Muller, G., Konig, H., Meissner, H., Toyka, K. V., & Rieckmann, P. (2006). "Fatigue" in multiple sclerosis. Development and and validation of the "Wurzburger Fatigue Inventory for MS". *Der Nervenarzt, 77*(2), 165-166, 168-170, 172-164.

Fleming, J. O., & Cook, T. D. (2006). Multiple sclerosis and the hygiene hypothesis. *Neurology, 67*(11), 2085-2086.

Flensner, G., Ek, A., & Söderhamn, O. (2003). Lived experience of MS-related fatigue-A phenomenological interview study. *International Journal of Nursing Studies, 40*(7), 707-717.

Ford, D., Jones, K., Middleton, R., Lockhart-Jones, H., Maramba, I., Noble, G.et al. (2012). The feasibility of collecting information from people with multiple sclerosis for the UK MS Register via a web portal: characterising a cohort of people with MS. *BMC Medical Informatics and Decision Making, 12*(1), 73.

Freal, J. E., Kraft, G. H., & Coryell, J. K. (1984). Symptomatic fatigue in multiple sclerosis. *Archives of Physical Medicine and Rehabilitation, 65*(3), 135-138.

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion, 14*(4), 577-606.

Fries, E., Dettenborn, L., & Kirschbaum, C. (2009). The cortisol awakening response (CAR): Facts and future directions. *International Journal of Psychophysiology, 72*(1), 67-73.

Fries, E., Hesse, J., Hellhammer, J., & Hellhammer, D. H. (2005). A new view on hypocortisolism. *Psychoneuroendocrinology, 30*(10), 1010-1016.

Frodl, T., & O'Keane, V. (2013). How does the brain deal with cumulative stress? A review with focus on developmental stress, HPA axis function and hippocampal structure in humans. *Neurobiology of Disease, 52*(0), 24-37.

Fukuda, K., Straus, S., Hickie, I., Sharpe, M., Dobbins, J., & Komaroff, A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Annals of Internal Medicine, 121*(12), 953-959.

Gaab, J., Huster, D., Peisen, R., Engert, V., Schad, T., Schurmeyer, T. H.et al. (2002). Low-dose dexamethasone suppression test in chronic fatigue syndrome and health. *Psychosomatic Medicine, 64*(2), 311-318.

Gallagher, P., Leitch, M. M., Massey, A. E., McAllister-Williams, R. H., & Young, A. H. (2006). Assessing cortisol and dehydroepiandrosterone (DHEA) in saliva: effects of collection method. *Journal of Psychopharmacology, 20*(5), 643-649.

Garde, A. H., & Hansen, A. M. (2005). Long-term stability of salivary cortisol. *Scandinavian Journal of Clinical & Laboratory Investigation, 65*(5), 433-436.

Gibbons, R. D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H. C., Greenhouse, J. B.et al. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data: Application to the nimh treatment of depression collaborative research program dataset. *Archives of General Psychiatry, 50*(9), 739-750.

Gil, K. M., Carson, J. W., Porter, L. S., Ready, J., Valrie, C., Redding-Lallinger, R.et al. (2003). Daily stress and mood and their association with pain, health-care use, and school activity in adolescents with sickle cell disease. *Journal of Pediatric Psychology, 28*(5), 363-373.

Gilchrest, B. A. (2008). Sun exposure and vitamin D sufficiency. *The American Journal of Clinical Nutrition, 88*(2), 570S-577S.

Giovannoni, G., Thompson, A. J., Miller, D. H., & Thompson, E. J. (2001). Fatigue is not associated with raised inflammatory markers in multiple sclerosis. *Neurology, 57*(4), 676-681.

References

Girden, E. R. (1992). *ANOVA Repeated Measures*. Newbury Park, CA: Sage Publications.

Gledhill, J. (2005). A qualitative study of the characteristics and representation of fatigue in a French speaking population of cancer patients and healthy subjects. *European Journal of Oncology Nursing, 9*(4), 294-312.

Gold, S. M., Kern, K. C., O'Connor, M.-F., Montag, M. J., Kim, A., Yoo, Y. S.et al. (2010). Smaller cornu ammonis 2-3/dentate gyrus volumes and elevated cortisol in multiple sclerosis patients with depressive symptoms. *Biological Psychiatry, 68*(6), 553-559.

Gold, S. M., Kruger, S., Ziegler, K. J., Krieger, T., Schulz, K. H., Otte, C.et al. (2011). Endocrine and immune substrates of depressive symptoms and fatigue in multiple sclerosis patients with comorbid major depression. *Journal of Neurology, Neurosurgery & Psychiatry, 82(7)*, 814-818.

Gold, S. M., Raji, A., Huitinga, I., Wiedemann, K., Schulz, K.-H., & Heesen, C. (2005). Hypothalamo-pituitary-adrenal axis activity predicts disease progression in multiple sclerosis. *Journal of Neuroimmunology, 165*(1-2), 186-191.

Goldman Consensus Group. (2005). The Goldman Consensus Statement on depression in multiple sclerosis. *Multiple Sclerosis, 11*(3), 328-337.

Goldman, M. D., Motl, R. W., & Rudick, R. A. (2010). Possible clinical outcome measures for clinical trials in patients with multiple sclerosis. *Therapeutic Advances in Neurological Disorders, 3*(4), 229-239.

Gorter, S. L., Bijlsma, J. W., Cutolo, M., Gomez-Reino, J., Kouloumas, M., Smolen, J. S.et al. (2010). Current evidence for the management of rheumatoid arthritis with glucocorticoids: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. *Annals of the Rheumatic Diseases, 69*(6), 1010-1014.

Gottschalk, M., Kümpfel, T., Flachenecker, P., Uhr, M., Trenkwalder, C., Holsboer, F.et al. (2005). Fatigue and regulation of the hypothalamo-pituitary-adrenal axis in multiple sclerosis. *Archives of Neurology, 62*(2), 277-280.

Gow, R., Thomson, S., Rieder, M., Van Uum, S., & Koren, G. (2010). An assessment of cortisol analysis in hair and its clinical applications. *Forensic Science International, 196*(1), 32-37.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549-576.

Grasser, A., Möller, A., Backmund, H., Yassouridis, A., & Holsboer, F. (1996). Heterogeneity of hypothalamic-pituitary-adrenal system response to a combined dexamethasone-CRH test in multiple sclerosis. *Experimental and Clinical Endocrinology & Diabetes, 104*(1), 31-37.

Gribbin, C. E., Watamura, S. E., Cairns, A., Harsh, J. R., & LeBourgeois, M. K. (2012). The cortisol awakening response (CAR) in 2- to 4-year-old children: Effects of acute nighttime sleep restriction, wake time, and daytime napping. *Developmental Psychobiology, 54*(4), 412-422.

Griefahn, B., & Robens, S. (2010). The normalization of the cortisol awakening response and of the cortisol shift profile across consecutive night shifts—An experimental study. *Psychoneuroendocrinology, 35*(10), 1501-1509.

Griefahn, B., & Robens, S. (2011). Cortisol awakening response - Are sampling delays of 15 minutes acceptable? *International Journal of Psychophysiology, 82*(2), 202-205.

Grilo, R. M., Treves, R., Preux, P. M., Vergne-Salle, P., & Bertin, P. (2007). Clinically relevant VAS pain score change in patients with acute rheumatic conditions. *Joint Bone Spine, 74*(4), 358-361.

Hacker, E. D., & Ferrans, C. E. (2007). Ecological momentary assessment of fatigue in patients receiving intensive cancer therapy. *Journal of Pain and Symptom Management, 33*(3), 267-275.

Hadjimichael, O., Vollmer, T., & Oleen-Burkey, M. (2008). Fatigue characteristics in multiple sclerosis: the North American Research Committee on Multiple Sclerosis (NARCOMS) survey. *Health and Quality of Life Outcomes, 6*, 100.

Hägg, O., Fritzell, P., & Nordwall, A. (2003). The clinical importance of changes in outcome scores after treatment for chronic low back pain. *European Spine Journal, 12*(1), 12-20.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*(1), 56-62.

Handel, A. E., Handunnetthi, L., Giovannoni, G., Ebers, G. C., & Ramagopalan, S. V. (2010). Genetic and environmental factors and the distribution of multiple sclerosis in Europe. *European Journal of Neurology, 17*(9), 1210-1214.

References

Hansen, Å. M., Garde, A. H., & Persson, R. (2008). Sources of biological and methodological variation in salivary cortisol and their impact on measurement among healthy adults: A review. *Scandinavian Journal of Clinical & Laboratory Investigation, 68*(6), 448-458.

Hansen, Å. M., Thomsen, J. F., Kaergaard, A., Kolstad, H. A., Kaerlev, L., Mors, O.et al. (2012). Salivary cortisol and sleep problems among civil servants. *Psychoneuroendocrinology, 37*(7), 1086-1095.

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The rand 36-item health survey 1.0. *Health Economics, 2*(3), 217-227.

Hays, R. D., & Stewart, A. L. (1992). Sleep measures. In A. L. Stewart & J. E. Ware (Eds.), *Measuring Functioning and Well-Being: the Medical Outcomes Study approach* Durham, NC: Duke University Press.

Heaney, J. L. J., Phillips, A. C., & Carroll, D. (2010). Ageing, depression, anxiety, social support and the diurnal rhythm and awakening response of salivary cortisol. *International Journal of Psychophysiology, 78*(3), 201-208.

Heesen, C., Gold, S. M., Huitinga, I., & Reul, J. M. H. M. (2007). Stress and hypothalamic-pituitary-adrenal axis function in experimental autoimmune encephalomyelitis and multiple sclerosis--A review. *Psychoneuroendocrinology, 32*(6), 604-618.

Heesen, C., Gold, S. M., Raji, A., Wiedemann, K., & Schulz, K.-H. (2002). Cognitive impairment correlates with hypothalamo–pituitary–adrenal axis dysregulation in multiple sclerosis. *Psychoneuroendocrinology, 27*(4), 505-517.

Heesen, C., Koehler, G., Gross, R., Tessmer, W., Schulz, K.-H., & Gold, S. M. (2005). Altered cytokine responses to cognitive stress in multiple sclerosis patients with fatigue. *Multiple Sclerosis, 11*(1), 51-57.

Heesen, C., Mohr, D. C., Huitinga, I., Bergh, F. T., Gaab, J., Otte, C.et al. (2007). Stress regulation in multiple sclerosis: current issues and concepts. *Multiple Sclerosis, 13*(2), 143-148.

Heesen, C., Nawrath, L., Reich, C., Bauer, N., Schulz, K., & Gold, S. M. (2006). Fatigue in multiple sclerosis: an example of cytokine mediated sickness behaviour? *Journal of Neurology, Neurosurgery & Psychiatry, 77*(1), 34-39.

Heesen, C., Schulz, H., Schmidt, M., Gold, S., Tessmer, W., & Schulz, K.-H. (2002). Endocrine and cytokine responses to acute psychological stress in multiple sclerosis. *Brain, Behavior, and Immunity, 16*(3), 282-287.

Heim, C., Ehlert, U., & Hellhammer, D. H. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology, 25*(1), 1-35.

Heim, C., Nater, U. M., Maloney, E., Boneva, R. S., Jones, J. F., & Reeves, W. C. (2009). Childhood trauma and risk for chronic fatigue syndrome: Association with neuroendocrine dysfunction. *Archives of General Psychiatry, 66*(1), 72-80.

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology, 34*(2), 163-171.

Hellhammer, J., Fries, E., Schweisthal, O. W., Schlotz, W., Stone, A. A., & Hagemann, D. (2007). Several daily measurements are necessary to reliably assess the cortisol rise after awakening: state- and trait components. *Psychoneuroendocrinology, 32*(1), 80-86.

Heuser, I., Yassouridis, A., & Holsboer, F. (1994). The combined dexamethasone/CRH test: a refined laboratory test for psychiatric disorders. *Journal of Psychiatric Research, 28*(4), 341-356.

Hewlett, S., Cockshott, Z., Byron, M., Kitchen, K., Tipler, S., Pope, D.et al. (2005). Patients' perceptions of fatigue in rheumatoid arthritis: Overwhelming, uncontrollable, ignored. *Arthritis Care & Research, 53*(5), 697-702.

Hirst, C., Swingler, R., Compston, D. A. S., Ben-Shlomo, Y., & Robertson, N. P. (2008). Survival and cause of death in multiple sclerosis: a prospective population-based study. *Journal of Neurology, Neurosurgery & Psychiatry, 79*(9), 1016-1021.

Hjollund, N., Andersen, J., & Bech, P. (2007). Assessment of fatigue in chronic disease: a bibliographic study of fatigue measurement scales. *Health and Quality of Life Outcomes, 5*(1), 1-5.

Horowitz, T. S., Cade, B. E., Wolfe, J. M., & Czeisler, C. A. (2001). Efficacy of bright light and sleep/darkness scheduling in alleviating circadian maladaptation to night work. *American Journal of Physiology - Endocrinology and Metabolism, 281*(2), E384-E391.

References

Horton, M., Rudick, R. A., Hara-Cleaver, C., & Marrie, R. A. (2010). Validation of a self-report comorbidity questionnaire for multiple sclerosis. *Neuroepidemiology, 35*(2), 83-90.

Houtveen, J. H., & Oei, N. Y. L. (2007). Recall bias in reporting medically unexplained symptoms comes from semantic memory. *Journal of Psychosomatic Research, 62*(3), 277-282.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Hox, J. J. (2010). *Multilevel analysis: techniques and applications* (Second Edition ed.). New York, NY: Routledge.

Hruschka, D. J., Kohrt, B. A., & Worthman, C. M. (2005). Estimating between- and within-individual variation in cortisol levels using multilevel models. *Psychoneuroendocrinology, 30*(7), 698-714.

Hufford, M. R. (2008). Special methodological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza & L. Nebeling (Eds.), *The Science of Real-Time Data Capture*. New York, NY: Oxford University Press.

Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. *Psychology of Addictive Behaviors, 16*(3), 205-211.

Inder, W. J., Prickett, T. C. R., & Mulder, R. T. (2005). Normal opioid tone and hypothalamic–pituitary–adrenal axis function in chronic fatigue syndrome despite marked functional impairment. *Clinical Endocrinology, 62*(3), 343-348.

Induruwa, I., Constantinescu, C. S., & Gran, B. (2012). Fatigue in multiple sclerosis—A brief review. *Journal of the Neurological Sciences, 323*(1-2), 9-15.

Ising, M., Horstmann, S., Kloiber, S., Lucae, S., Binder, E. B., Kern, N.et al. (2007). Combined dexamethasone/corticotropin releasing hormone test predicts treatment response in major depression - a potential biomarker? *Biological Psychiatry, 62*(1), 47-54.

Jacobs, N., Myin-Germeys, I., Derom, C., Delespaul, P., van Os, J., & Nicolson, N. A. (2007). A momentary assessment study of the relationship between affective and adrenocortical stress responses in daily life. *Biological Psychology, 74*(1), 60-66.

Jamison, R. N., Gracely, R. H., Raymond, S. A., Levine, J. G., Marino, B., Herrmann, T. J.et al. (2002). Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers. *Pain, 99*(1–2), 341-347.

Janardhan, V., & Bakshi, R. (2002). Quality of life in patients with multiple sclerosis: The impact of fatigue and depression. *Journal of the Neurological Sciences, 205*(1), 51-58.

Janssens, A. C. J. W., van Doorn, P. A., de Boer, J. B., van der Meché, F. G. A., Passchier, J., & Hintzen, R. Q. (2003). Impact of recently diagnosed multiple sclerosis on quality of life, anxiety, depression and distress of patients and partners. *Acta Neurologica Scandinavica, 108*(6), 389-395.

Jason, L. A., Corradi, K., Torres-Harding, S., Taylor, R. R., & King, C. (2005). Chronic fatigue syndrome: The need for subtypes. *Neuropsychology Review, 15*(1), 29-58.

Jensen, M. P., Mardekian, J., Lakshminarayanan, M., & Boye, M. E. (2008). Validity of 24-h recall ratings of pain severity: Biasing effects of "peak" and "end" pain. *Pain, 137*(2), 422-427.

Jerjes, W. K., Cleare, A. J., Wessely, S., Wood, P. J., & Taylor, N. F. (2005). Diurnal patterns of salivary cortisol and cortisone output in chronic fatigue syndrome. *Journal of Affective Disorders, 87*(2-3), 299-304.

Johansson, S., Ytterberg, C., Gottberg, K., Widén Holmqvist, L., & von Koch, L. (2009). Use of health services in people with multiple sclerosis with and without fatigue. *Multiple Sclerosis, 15*(1), 88-95.

Johansson, S., Ytterberg, C., Hillert, J., Holmqvist, L. W., & von Koch, L. (2008). A longitudinal study of variations in and predictors of fatigue in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 79*(4), 454-457.

Johnson, K. L., Yorkston, K. M., Klasner, E. R., Kuehn, C. M., Johnson, E., & Amtmann, D. (2004). The cost and benefits of employment: a qualitative study of experiences of persons with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation, 85*(2), 201-209.

Jones, K. H., Ford, D. V., Jones, P. A., John, A., Middleton, R. M., Lockhart-Jones, H.et al. (2012). A large-scale study of anxiety and depression in people with multiple sclerosis: A survey via the web portal of the UK MS Register. *Plos One, 7*(7), e41910-e41910.

References

Jung, C., Ho, J. T., Torpy, D. J., Rogers, A., Doogue, M., Lewis, J. G.et al. (2011). A longitudinal study of plasma and urinary cortisol in pregnancy and postpartum. *The Journal of Clinical Endocrinology & Metabolism, 96*(5), 1533-1540.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4*(6), 401-405.

Kaiser, H. F. (1974). Analysis of factorial simplicity. *Psychometrika, 39*, 31-36.

Kaminska, M., Kimoff, R. J., Schwartzman, K., & Trojan, D. A. (2011). Sleep disorders and fatigue in multiple sclerosis: Evidence for association and interaction. *Journal of the Neurological Sciences, 302*(1–2), 7-13.

Kamm, C. P., Heldner, M. R., Vanbellingen, T., Mattle, H. P., Müri, R., & Bohlhalter, S. (2012). Limb apraxia in multiple sclerosis: Prevalence and impact on manual dexterity and activities of daily living. *Archives of Physical Medicine and Rehabilitation, 93*(6), 1081-1085.

Kanchandani, R., & Howe, J. G. (1982). Lhermitte's sign in multiple sclerosis: a clinical survey and review of the literature. *Journal of Neurology, Neurosurgery and Psychiatry, 45*(4), 308-312.

Kang, J. H., Chen, Y. H., & Lin, H. C. (2010). Comorbidities amongst patients with multiple sclerosis: A population-based controlled study. *European Journal of Neurology, 17*(9), 1215-1219.

Kappos, L., Freedman, M. S., Polman, C. H., Edan, G., Hartung, H.-P., Miller, D. H.et al. (2007). Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: a 3-year follow-up analysis of the BENEFIT study. *The Lancet, 370*(9585), 389-397.

Kern, S., Krause, I., Horntrich, A., Thomas, K., Aderhold, J., & Ziemssen, T. (2013). Cortisol awakening response is linked to disease course and progression in multiple sclerosis. *Plos One, 8*(4), e60647.

Kern, S., Schultheiß, T., Schneider, H., Schrempf, W., Reichmann, H., & Ziemssen, T. (2011). Circadian cortisol, depressive symptoms and neurological impairment in early multiple sclerosis. *Psychoneuroendocrinology, 36*(10), 1505-1512.

Kerns, R. D., Turk, D. C., & Rudy, T. E. (1985). The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain, 23*(4), 345-356.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R.et al. (2003). The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (ncs-r). *JAMA, 289*(23), 3095-3105.

Khani, S., & Tayek, J. A. (2001). Cortisol increases gluconeogenesis in humans: its role in the metabolic syndrome. *Clinical Science, 101*(6), 739-747.

Kiecolt-Glaser, J. K., Glaser, R., Shuttleworth, E. C., Dyer, C. S., Ogrocki, P., & Speicher, C. E. (1987). Chronic stress and immunity in family caregivers of Alzheimer's disease victims. *Psychosomatic Medicine, 49*(5), 523-535.

Kim, E., Lovera, J., Schaben, L., Melara, J., Bourdette, D., & Whitham, R. (2010). Novel method for measurement of fatigue in multiple sclerosis: real-time digital fatigue score. *Journal of Rehabilitation Research and Development, 47*(5), 477-484.

Kim, J., Kikuchi, H., & Yamamoto, Y. (2013). Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. *British Journal of Health Psychology, 18*(1), 155-167.

Kirschbaum, C., & Hellhammer, D. H. (1989). Salivary cortisol in psychobiological research: An overview. *Neuropsychobiology, 22*(3), 150-169.

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine, 61*(2), 154-162.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1-2), 76-81.

Kirschbaum, C., Wüst, S., & Hellhammer, D. (1992). Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine, 54*(6), 648-657.

Kister, I., Chamot, E., Salter, A. R., Cutter, G. R., Bacon, T. E., & Herbert, J. (2013). Disability in multiple sclerosis: A reference for patients and clinicians. *Neurology, 80*(11), 1018-1024.

Knorr, U., Vinberg, M., Kessing, L. V., & Wetterslev, J. (2010). Salivary cortisol in depressed patients versus control persons: A systematic review and meta-analysis. *Psychoneuroendocrinology, 35*(9), 1275-1286.

References

Kocalevent, R., Hinz, A., Brahler, E., & Klapp, B. (2011). Determinants of fatigue and stress. *BMC Research Notes, 4*(1), 238.

Koch-Henriksen, N., & Sørensen, P. S. (2010). The changing demographic pattern of multiple sclerosis epidemiology. *The Lancet Neurology, 9*(5), 520-532.

Koch, M., Uyttenboogaart, M., van Harten, A., & De Keyser, J. (2008). Factors associated with the risk of secondary progression in multiple sclerosis. *Multiple Sclerosis, 14*(6), 799-803.

Korn, T. (2008). Pathophysiology of multiple sclerosis. *Journal of Neurology, 255*(0), 2-6.

Kos, D., Kerckhofs, E., Nagels, G., D'hooghe, M. B., & Ilsbroukx, S. (2008). Origin of fatigue in multiple sclerosis: Review of the literature. *Neurorehabilitation and Neural Repair, 22*(1), 91-100.

Kroencke, D. C., Lynch, S. G., & Denney, D. R. (2000). Fatigue in multiple sclerosis: relationship to depression, disability, and disease pattern. *Multiple Sclerosis, 6*(2), 131-136.

Krupp, L. B., Alvarez, L. A., LaRocca, N. G., & Scheinberg, L. C. (1988). Fatigue in multiple sclerosis. *Archives of Neurology, 45*(4), 435-437.

Krupp, L. B., Coyle, P. K., Doscher, C., Miller, A., Cross, A. H., Jandorf, L.et al. (1995). Fatigue therapy in multiple sclerosis: Results of a double-blind, randomized, parallel trial of amantadine, pemoline, and placebo. *Neurology, 45*(11), 1956-1961.

Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The Fatigue Severity Scale: Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology, 46*(10), 1121-1123.

Krupp, L. B., Serafin, D. J., & Christodoulou, C. (2010). Multiple sclerosis-associated fatigue. *Expert Review of Neurotherapeutics, 10*(9), 1437-1447.

Kudielka, B. M., Broderick, J. E., & Kirschbaum, C. (2003). Compliance with saliva sampling protocols: electronic monitoring reveals invalid cortisol daytime profiles in noncompliant subjects. *Psychosomatic Medicine, 65*(2), 313-319.

Kudielka, B. M., Gierens, A., Hellhammer, D. H., Wüst, S., & Schlotz, W. (2012). Salivary cortisol in ambulatory assessment - Some dos, some don'ts, and some open questions. *Psychosomatic Medicine, 74*(4), 418-431.

Kumari, M., Badrick, E., Chandola, T., Adam, E. K., Stafford, M., Marmot, M. G.et al. (2009). Cortisol secretion and fatigue: associations in a community based cohort. *Psychoneuroendocrinology, 34*(10), 1476-1485.

Kunz-Ebrecht, S. R., Kirschbaum, C., Marmot, M., & Steptoe, A. (2004). Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. *Psychoneuroendocrinology, 29*(4), 516-528.

Kurina, L. M., Schneider, B., & Waite, L. J. (2004). Stress, symptoms of depression and anxiety, and cortisol patterns in working parents. *Stress and Health, 20*(2), 53-63.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis. *Neurology, 33*(11), 1444-1452.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963-974.

Larsson, C., Gullberg, B., Rastam, L., & Lindblad, U. (2009). Salivary cortisol differs with age and sex and shows inverse associations with WHR in Swedish women: a cross-sectional study. *BMC Endocrine Disorders, 9*(1), 16.

Lassmann, H., van Horssen, J., & Mahad, D. (2012). Progressive multiple sclerosis: pathology and pathogenesis. *Nature Reviews Neurology, 8*(11), 647-656.

Law, R., Hucklebridge, F., Thorn, L., Evans, P., & Clow, A. (2013). State variation in the cortisol awakening response. *Stress, 16*(5), 483-492.

Lazarus, R. S., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York: Springer.

Lemmens, P., Knibbe, R. A., & Tan, F. (1988). Weekly recall and dairy estimates of alcohol consumption in a general population survey. *Journal of Studies on Alcohol, 49*(2), 131-135.

LePine, J. A., LePine, M. A., & Jackson, C. L. (2004). Challenge and hindrance stress: Relationships with exhaustion, motivation to learn, and learning performance. *Journal of Applied Psychology, 89*(5), 883-891.

Lerdal, A., Celius, E. G., Krupp, L. B., & Dahl, A. A. (2007). A prospective study of patterns of fatigue in multiple sclerosis. *European Journal of Neurology, 14*(12), 1338-1343.

References

Lerdal, A., Celius, E. G., & Moum, T. (2003). Fatigue and its association with sociodemographic variables among multiple sclerosis patients. *Multiple Sclerosis, 9*(5), 509-514.

Lerdal, A., Wahl, A. K., Rustoen, T., Hanestad, B. R., & Moum, T. (2005). Fatigue in the general population: A translation and test of the psychometric properties of the Norwegian version of the fatigue severity scale. *Scandinavian Journal of Public Health, 33*(2), 123-130.

Levine, A., Zagoory-Sharon, O., Feldman, R., Lewis, J. G., & Weller, A. (2007). Measuring cortisol in human psychobiological studies. *Physiology & Behavior, 90*(1), 43-53.

Levine, S. (2005). Developmental determinants of sensitivity and resistance to stress. *Psychoneuroendocrinology, 30*(10), 939-946.

Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992). Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychological Medicine, 22*(2), 465-486.

Liddle, G. W. (1960). Tests of pituitary-adrenal suppressibility in the diagnosis of Cushing's syndrome. *The Journal of Clinical Endocrinology & Metabolism, 20*(12), 1539-1560.

Lindeberg, S. I., Eek, F., Lindbladh, E., Östergren, P.-O., Hansen, Å. M., & Karlson, B. (2008). Exhaustion measured by the SF-36 vitality scale is associated with a flattened diurnal cortisol profile. *Psychoneuroendocrinology, 33*(4), 471-477.

Liu, S., Rovine, M. J., & Molenaar, P. C. M. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods, 17*(1), 15-30.

Liu, X. J., Ye, H. X., Li, W. P., Dai, R., Chen, D., & Jin, M. (2009). Relationship between psychosocial factors and onset of multiple sclerosis. *European Neurology, 62*(3), 130-136.

Maas, C. J. M., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis, 46*(3), 427-440.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19-40.

Mackenzie, I. S., Morant, S. V., Bloomfield, G. A., MacDonald, T. M., & O'Riordan, J. (2014). Incidence and prevalence of multiple sclerosis in the UK 1990–2010: a descriptive study in the General Practice Research Database. *Journal of Neurology, Neurosurgery & Psychiatry, 85*(1), 76-84.

Mainero, C., Faroni, J., Gasperini, C., Filippi, M., Giugni, E., Ciccarelli, O.et al. (1999). Fatigue and magnetic resonance imaging activity in multiple sclerosis. *Journal of Neurology, 246*(6), 454-458.

Malcomson, K. S., Lowe-Strong, A. S., & Dunwoody, L. (2008). What can we learn from the personal insights of individuals living and coping with multiple sclerosis? *Disability and Rehabilitation, 30*(9), 662-674.

Manconi, M., Ferini-Strambi, L., Filippi, M., Bonanni, E., Iudice, A., Murri, L.et al. (2008). Multicenter case-control study on restless legs syndrome in multiple sclerosis: the REMS study. *Sleep, 31*(7), 944-952.

Marmot, M., & Brunner, E. (2005). Cohort profile: The Whitehall II study. *International Journal of Epidemiology, 34*(2), 251-256.

Marrie, R. A., Fisher, E., Miller, D. M., Lee, J.-C., & Rudick, R. A. (2005). Association of fatigue and brain atrophy in multiple sclerosis. *Journal of the Neurological Sciences, 228*(2), 161-166.

Marrie, R. A., Horwitz, R., Cutter, G., Tyry, T., Campagnolo, D., & Vollmer, T. (2008). Comorbidity, socioeconomic status and multiple sclerosis. *Multiple Sclerosis, 14*(8), 1091-1098.

Mason, J. W., Hartley, L. H., Kotchen, T. A., Mougey, E. H., Ricketts, P. T., & Jones, L. G. (1973). Plasma cortisol and norepinephrine responses in anticipation of muscular exercise. *Psychosomatic Medicine, 35*(5), 406-414.

Matlab (Version 7.10.0) [Computer Software]. Natick, MA: The Mathsworks Inc.

McCabe, K. O., Mack, L., & Fleeson, W. (2011). A guide for data cleaning in experience sampling studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Press.

McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D.et al. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology, 50*(1), 121-127.

References

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *The New England Journal Of Medicine, 338*(3), 171-179.

McEwen, B. S. (2004). Protection and damage from acute and chronic stress: Allostasis and allostatic overload and relevance to the pathophysiology of psychiatric disorders. *Annals of The New York Academy Of Sciences, 1032*(1), 1-7.

McEwen, B. S., & Seeman, T. (1999). Protective and damaging effects of mediators of stress: Elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences, 896*(1), 30-47.

McFarland, H. F., & Martin, R. (2007). Multiple sclerosis: a complicated picture of autoimmunity. [10.1038/ni1507]. *Nature Immunology, 8*(9), 913-919.

McGough, J. J., & Faraone, S. V. (2009). Estimating the size of treatment effects: moving beyond p values. *Psychiatry, 6*(10), 21-29.

McGrath, R. E. (2006). When effect sizes disagree: The case or r and d. *Psychological Methods, 11*(4), 386-401.

McHorney, C. A. (2000). Concepts and Measurement of Health Status and Health-related Quality of Life. In G. L. Abrecht, R. Fitzpatrick & S. C. Scrimshaw (Eds.), *The Handbook of Social Studies in Health and Medicine*. London, UK: SAGE Publications.

McKenzie, R., O'Fallon, A., Dale, J., Demitrack, M., Sharma, G., Deloria, M.et al. (1998). Low-dose hydrocortisone for treatment of chronic fatigue syndrome: a randomized controlled trial. *JAMA, 280*(12), 1061-1066.

McLean, S. A., Williams, D. A., Harris, R. E., Kop, W. J., Groner, K. H., Ambrose, K.et al. (2005). Momentary relationship between cortisol secretion and symptoms in patients with fibromyalgia. *Arthritis & Rheumatism, 52*(11), 3660-3669.

McNair, D. M., Lorr, M., & Dropleman, L. (1971). *EITS Manual for the Profile of Mood States*. San Diego, CA.: Educational and Industrial Testing Service.

Meijer, O. C., de Lange, E. C. M., Breimer, D. D., de Boer, A. G., Workel, J. O., & de Kloet, E. R. (1998). Penetration of dexamethasone into brain glucocorticoid targets is enhanced in mdr1A P-glycoprotein knockout mice. *Endocrinology, 139*(4), 1789-1793.

Melief, J., Wit, S., Eden, C., Teunissen, C., Hamann, J., Uitdehaag, B.et al. (2013). HPA axis activity in multiple sclerosis correlates with disease severity, lesion type and gene expression in normal-appearing white matter. *Acta Neuropathologica*, 1-13.

Mendoza, T. R., Wang, X. S., Cleeland, C. S., Morrissey, M., Johnson, B. A., Wendt, J. K.et al. (1999). The rapid assessment of fatigue severity in cancer patients. *Cancer, 85*(5), 1186-1196.

Michelson, D., Stone, L., Galliven, E., Magiakou, M. A., Chrousos, G. P., Sternberg, E. M.et al. (1994). Multiple sclerosis is associated with alterations in hypothalamic-pituitary-adrenal axis function. *The Journal of Clinical Endocrinology & Metabolism, 79*(3), 848-853.

Michielsen, H. J., Willemsen, T. M., Croon, M. A., de Vries, J., & Van Heck, G. L. (2004). Determinants of general fatigue and emotional exhaustion: a prospective study. *Psychology & Health, 19*(2), 223-235.

Miller, D. B., & O'Callaghan, J. P. (2002). Neuroendocrine aspects of the response to stress. *Metabolism: Clinical and Experimental, 51*(6), 5-10.

Miller, D. H., & Leary, S. M. (2007). Primary-progressive multiple sclerosis. *The Lancet Neurology, 6*(10), 903-912.

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin, 133*(1), 25-45.

Miller, G. E., Cohen, S., & Ritchey, A. K. (2002). Chronic psychological stress and the regulation of pro-inflammatory cytokines: A glucocorticoid-resistance model. *Health Psychology, 21*(6), 531-541.

Miller, R., Plessow, F., Kirschbaum, C., & Stalder, T. (2013). Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. *Psychosomatic Medicine, 75*(9), 832-840.

Milligan, N. M., Newcombe, R., & Compston, D. A. (1987). A double-blind controlled trial of high dose methylprednisolone in patients with multiple sclerosis: 1. Clinical effects. *Journal of Neurology, Neurosurgery & Psychiatry, 50*(5), 511-516.

Mills, R. J., & Young, C. A. (2008). A medical definition of fatigue in multiple sclerosis. *QJM, 101*(1), 49-60.

Mills, R. J., & Young, C. A. (2010). The relationship between fatigue and other clinical features of multiple sclerosis. *Multiple Sclerosis*.

References

Minden, S. L., Frankel, D., Hadden, L., Perloffp, J., Srinath, K. P., & Hoaglin, D. C. (2006). The Sonya Slifka Longitudinal Multiple Sclerosis Study: methods and sample characteristics. *Multiple Sclerosis, 12*(1), 24-38.

Miron-Shatz, T., Stone, A. A., & Kahneman, D. (2009). Memories of yesterday's emotions: does the valence of experience affect the memory-experience gap. *Emotion, 9*(6), 885-891.

Mohr, D. C., Goodkin, D. E., Bacchetti, P., Boudewyn, A. C., Huang, L., Marrietta, P.et al. (2000). Psychological stress and the subsequent appearance of new brain MRI lesions in MS. *Neurology, 55*(1), 55-61.

Mohr, D. C., Goodkin, D. E., Likosky, W., Beutler, L., Gatto, N., & Langan, M. K. (1997). Identification of Beck Depression Inventory items related to multiple sclerosis. *Journal of Behavioral Medicine, 20*(4), 407-414.

Mohr, D. C., Hart, S. L., Julian, L., Cox, D., & Pelletier, D. (2004). Association between stressful life events and exacerbation in multiple sclerosis: a meta-analysis. *BMJ: British Medical Journal, 328*(7442), 731-733.

Mohr, D. C., Lovera, J., Brown, T., Cohen, B., Neylan, T., Henry, R.et al. (2012). A randomized trial of stress management for the prevention of new brain lesions in MS. *Neurology, 79*(5), 412-419.

Mohr, D. C., & Pelletier, D. (2006). A temporal framework for understanding the effects of stressful life events on inflammation in patients with multiple sclerosis. *Brain, Behavior, and Immunity, 20*(1), 27-36.

Mollaoğlu, M., & Üstün, E. (2009). Fatigue in multiple sclerosis patients. *Journal of Clinical Nursing, 18*(9), 1231-1238.

Morris, M. E., Cantwell, C., Vowels, L., & Dodd, K. (2002). Changes in gait and fatigue from morning to afternoon in people with multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 72*(3), 361-365.

Morriss, R., Wearden, A., & Mullis, R. (1998). Exploring the validity of the chalder fatigue scale in chronic fatigue syndrome. *Journal of Psychosomatic Research, 45*(5), 411-417.

Multiple Sclerosis Council for Clinical Practice Guidelines. (1998). *Multiple Sclerosis Council for Clinical Practice Guidelines: Fatigue and Multiple Sclerosis*. Washington, DC: Paralyzed Veterans of America.

Munger, K. L., Zhang, S. M., O'Reilly, E., Hernán, M. A., Olek, M. J., Willett, W. C.et al. (2004). Vitamin D intake and incidence of multiple sclerosis. *Neurology, 62*(1), 60-65.

Murray, T. J. (1985). Amantadine therapy for fatigue in multiple sclerosis. *The Canadian Journal of Neurological Sciences, 12*(3), 251-254.

Murray, T. J. (2006). Diagnosis and treatment of multiple sclerosis. *British Medical Journal, 332*(7540), 525-527.

Naggara, O., Raymond, J., Guilbert, F., Roy, D., Weill, A., & Altman, D. (2011). Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. *American Journal of Neuroradiology, 32*(3), 437-440.

Nater, U. M., Maloney, E., Boneva, R. S., Gurbaxani, B. M., Lin, J.-M., Jones, J. F.et al. (2008). Attenuated morning salivary cortisol concentrations in a population-based study of persons with chronic fatigue syndrome and well controls. *The Journal of Clinical Endocrinology & Metabolism, 93*(3), 703-709.

Neema, M., Caccarelli, A., Jackson, J. S., & Bakshi, R. (2012). Magnetic resonance imaging in multiple sclerosis. In H. L. Weiner & J. M. Stankiewicz (Eds.), *Multiple Sclerosis: Diagnosis and Therapy*. Chichester, UK: Wiley-Blackwell.

Nezlek, J. B. (2007). A multilevel framework for understanding relationships among traits, states, situations and behaviours. *European Journal of Personality, 21*(6), 789-810.

Nezlek, J. B. (2011). Multilevel modeling analyses of diary-style data. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Press.

NICE. (2003). *Multiple Sclerosis: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care*. London, UK: National Institute for Clinical Excellence.

NICE. (2010). *Natalizumab for the treatment of adults with highly active relapsing-remitting multiple sclerosis*. London, UK: National Institute for Health and Clinical Excellence.

NICE. (2013). ESUOM9 Fatigue in Multiple Sclerosis: Modafinil [online] Retrieved December 19, 2013, from http://www.nice.org.uk/mpc/evidencesummariesunlicensedofflabelmedicines/ESUOM9.jsp

Nicolson, N. A. (2008). Measurement of cortisol. In L. J. Luecken & L. G. Gallo (Eds.), *Handbook of Physiological Research Methods in Health Psychology*. Thousand Oaks, CA: SAGE Publications.

References

Noonan, C. W., Kathman, S. J., & White, M. C. (2002). Prevalence estimates for MS in the United States and evidence of an increasing trend for women. *Neurology, 58*(1), 136-138.

Noseworthy, J. H., Lucchinetti, C., Rodriguez, M., & Weinshenker, B. G. (2000). Multiple sclerosis. *The New England Journal Of Medicine, 343*(13), 938-952.

O'Hara, R., Schröder, C. M., Mahadevan, R., Schatzberg, A. F., Lindley, S., Fox, S.et al. (2007). Serotonin transporter polymorphism, memory and hippocampal volume in the elderly: association and interaction with cortisol. *Molecular Psychiatry, 12*(6), 544-555.

Office for National Statistics. (2013, January 15 2014). Full Report - Graduates in the UK Labour Market [online] Retrieved January 15, 2014, from http://www.ons.gov.uk/ons/rel/lmac/graduates-in-the-labour-market/2013/rpt---graduates-in-the-uk-labour-market-2013.html

Okun, M. L., Krafty, R. T., Buysse, D. J., Monk, T. H., Reynolds Iii, C. F., Begley, A.et al. (2010). What constitutes too long of a delay? Determining the cortisol awakening response (CAR) using self-report and PSG-assessed wake time. *Psychoneuroendocrinology, 35*(3), 460-468.

Olsson, M., Lexell, J., & Söderberg, S. (2005). The meaning of fatigue for women with multiple sclerosis. *Journal of Advanced Nursing, 49*(1), 7-15.

Otte, C., Hart, S., Neylan, T. C., Marmar, C. R., Yaffe, K., & Mohr, D. C. (2005). A meta-analysis of cortisol response to challenge in human aging: importance of gender. *Psychoneuroendocrinology, 30*(1), 80-91.

Oullet, M., & Morin, C. (2006). Fatigue following traumatic brain injury: Frequency, characteristics, and associated factors. *Rehabilitation Psychology, 51*(2), 140-149.

Pakpoor, J., Giovannoni, G., & Ramagopalan, S. V. (2013). Epstein–Barr virus and multiple sclerosis: association or causation? *Expert Review of Neurotherapeutics, 13*(3), 287-297.

Papadopoulos, A., & Cleare, A. J. (2012). Hypothalamic-pituitary-adrenal axis dysfunction in chronic fatigue syndrome. *Nature Reviews Endocrinology, 8*(1), 22-32.

Papadopoulos, A., Ebrecht, M., Roberts, A. D. L., Poon, L., Rohleder, N., & Cleare, A. J. (2009). Glucocorticoid receptor mediated negative feedback in chronic fatigue syndrome using the low dose (0.5 mg) dexamethasone suppression test. *Journal of Affective Disorders, 112*(1-3), 289-294.

Pariante, C. M., & Lightman, S. L. (2008). The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences, 31*(9), 464-468.

Patrick, E., Christodoulou, C., & Krupp, L. B. (2009). Longitudinal correlates of fatigue in multiple sclerosis. *Multiple Sclerosis, 15*(2), 258-261.

Patten, S. B., Beck, C. A., Williams, J. V., Barbui, C., & Metz, L. M. (2003). Major depression in multiple sclerosis: a population-based perspective. *Neurology, 61*(11), 1524-1527.

Pawlikowska, T., Chalder, T., Hirsch, S. R., Wallace, P., Wright, D. J. M., & Wessely, S. C. (1994). Population based study of fatigue and psychological distress. *British Medical Journal, 308*(6931), 763-766.

Peeters, F., Nicholson, N. A., & Berkhof, J. (2003). Cortisol responses to daily events in major depressive disorder. *Psychosomatic Medicine, 65*(5), 836-841.

Peters, M. L., Sorbi, M. J., Kruise, D. A., Kerssens, J. J., Verhaak, P. F. M., & Bensing, J. M. (2000). Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. *Pain, 84*(2–3), 181-192.

Petrowski, K., Paul, S., Albani, C., & Brahler, E. (2012). Factor structure and psychometric properties of the trier inventory for chronic stress (TICS) in a representative german sample. *BMC Medical Research Methodology, 12*(1), 42.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85-112.

Pittion-Vouyovitch, S., Debouverie, M., Guillemin, F., Vandenberghe, N., Anxionnat, R., & Vespignani, H. (2006). Fatigue in multiple sclerosis is related to disability, depression and quality of life. *Journal of the Neurological Sciences, 243*(1-2), 39-45.

Poikolainen, K. (1985). Underestimation of recalled alcohol intake in relation to actual consumption. *British Journal of Addiction, 80*(2), 215-216.

References

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M.et al. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology, 69*(2), 292-302.

Polman, C. H., Reingold, S. C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L.et al. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals of Neurology, 58*(6), 840-846.

Porter, L. S., Gil, K. M., Carson, J. W., Anthony, K. K., & Ready, J. (2000). The role of stress and mood in sickle cell disease pain. *Journal of Health Psychology, 5*(1), 53-63.

Poteliakhoff, A. (1981). Adrenocortical activity and some clinical findings in acute and chronic fatigue. *Journal of Psychosomatic Research, 25*(2), 91-95.

Powell, D. J., & Schlotz, W. (2012). Daily life stress and the cortisol awakening response: testing the anticipation hypothesis. *Plos One, 7*(12), e52067.

Prince, S., Adamo, K., Hamel, M., Hardt, J., Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 56.

PRISMS Study Group. (1998). Randomised double-blind placebo-controlled study of interferon β-1a in relapsing/remitting multiple sclerosis. *The Lancet, 352*(9139), 1498-1504.

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*(7), 916-931.

Pruessner, J. C., Wolf, O. T., Hellhammer, D. H., Buske-Kirschbaum, A., von Auer, K., Jobst, S.et al. (1997). Free cortisol levels after awakening: a reliable biological marker for the assessment of adrenocortical activity. *Life Sciences, 61*(26), 2539-2549.

Pruessner, M., Hellhammer, D. H., Pruessner, J. C., & Lupien, S. J. (2003). Self-reported depressive symptoms and stress levels in healthy young men: Associations with the cortisol response to awakening. *Psychosomatic Medicine, 65*(1), 92-99.

Pucci, E., Giuliani, G., Solari, A., Di Pietrantonj, C., Simi, S., Minozzi, S.et al. (2011). Natalizumab for relapsing remitting multiple sclerosis. *Cochrane Database of Systematic Reviews, 10*(CD007621).

Putzki, N., Katsarava, Z., Vago, S., Diener, H. C., & Limmroth, V. (2008). Prevalence and severity of multiple-sclerosis-associated fatigue in treated and untreated patients. *European Neurology, 59*(3-4), 136-142.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385-401.

Radloff, L. S. (1991). The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. *Journal of Youth and Adolescence, 20*(2), 149-166.

Rahman, K., Burton, A., Galbraith, S., Lloyd, A., & Vollmer-Conna, U. (2011). Sleep-wake behavior in chronic fatigue syndrome. *Sleep, 34*(5), 671-678.

Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology, 41*(5), 685-691.

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*(1), 1-17.

Reder, A. T., Lowy, M. T., Meltzer, H. Y., & Antel, J. P. (1987). Dexamethasone suppression test abnormalities in multiple sclerosis relation to ACTH therapy. *Neurology, 37*(5), 849-849.

Reis, H. T. (2011). Why Researchers Should Think "Real-World". In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily LifE*. New York, NY: Guilford Press.

Repping-Wuts, H., Uitterhoeve, R., van Riel, P., & van Achterberg, T. (2008). Fatigue as experienced by patients with rheumatoid arthritis (RA): A qualitative study. *International Journal of Nursing Studies, 45*(7), 995-1002.

Review Manager (Version 5.1) [Compute software]. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.

Ricci, J. A., Chee, E., Lorandeau, A. L., & Berger, J. (2007). Fatigue in the U.S. workforce: Prevalence and implications for lost productive work time. *Journal of Occupational and Environmental Medicine, 49*(1), 1-10.

Rice, C. M., Cottrell, D., Wilkins, A., & Scolding, N. J. (2013). Primary progressive multiple sclerosis: progress and challenges. *Journal of Neurology, Neurosurgery & Psychiatry, 84*(10), 1100-1106.

References

Richards, R. G., Sampson, F. C., Beard, S. M., & Tappenden, P. (2002). A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models. *Health Technology Assessment, 6*(10), 1-73.

Ritvo, P. G., Fischer, J. S., Miller, D. M., Andrews, H., Paty, D., & La Rocca, N. G. (1997). *Multiple Sclerosis Quality of Life Inventory: A User's Manual*. New York, NY.: National Multiple Sclerosis Society.

Roberts, A. D. L., Wessely, S., Chalder, T., Papadopoulos, A., & Cleare, A. J. (2004). Salivary cortisol response to awakening in chronic fatigue syndrome. *The British Journal of Psychiatry, 184*(2), 136-141.

Rohleder, N., & Kirschbaum, C. (2006). The hypothalamic–pituitary–adrenal (HPA) axis in habitual smokers. *International Journal of Psychophysiology, 59*(3), 236-243.

Roosendaal, S. D., Geurts, J. J. G., Vrenken, H., Hulst, H. E., Cover, K. S., Castelijns, J. A.et al. (2009). Regional DTI differences in multiple sclerosis patients. *NeuroImage, 44*(4), 1397-1403.

Rosenberg, J. H., & Shafor, R. (2005). Fatigue in multiple sclerosis: a rational approach to evaluation and treatment. *Current Neurology and Neuroscience Reports, 5*(2), 140-146.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Rosmond, R. (2003). Stress induced disturbances of the HPA axis: a pathway to Type 2 diabetes? *Medical Science Monitor, 9*(2), 35-39.

Rudick, R. A., Lee, J., Cutter, G. R., Miller, D. M., Bourdette, D., Weinstock-Guttman, B.et al. (2010). Disability progression in a clinical trial of relapsing-remitting multiple sclerosis: Eight-year follow-up. *Archives of Neurology, 67*(11), 1329-1335.

Russell, E., Koren, G., Rieder, M., & Van Uum, S. (2012). Hair cortisol as a biological marker of chronic stress: Current status, future directions and unanswered questions. *Psychoneuroendocrinology, 37*(5), 589-601.

Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology, 36*(3), 666-676.

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews, 21*(1), 55-89.

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition & Emotion, 15*(1), 81-97.

Schimmack, U. (2008). The structure of subjective well-being. In M. Eid & R. J. Larsen (Eds.), *The Science of Subjective Well-Being*. New York, NY: The Guilford Press.

Schlotz, W. (2011). Ambulatory Psychoneuroendocrinology: Assessing Salivary Cortisol and Other Hormones in Daily Life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Press.

Schlotz, W., Hellhammer, J., Schulz, P., & Stone, A. A. (2004). Perceived work overload and chronic worrying predict weekend-weekday differences in the cortisol awakening response. *Psychosomatic Medicine, 66*(2), 207-214.

Schlotz, W., Kumsta, R., Layes, I., Entringer, S., Jones, A., & Wüst, S. (2008). Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: A question of timing. *Psychosomatic Medicine, 70*(7), 787-796.

Schlotz, W., Schulz, P., Hellhammer, J., Stone, A. A., & Hellhammer, D. H. (2006). Trait anxiety moderates the impact of performance pressure on salivary cortisol in everyday life. *Psychoneuroendocrinology, 31*(4), 459-472.

Schneider, S., Stone, A. A., Schwartz, J. E., & Broderick, J. E. (2011). Peak and end effects in patients' daily recall of pain and fatigue: A within-subjects analysis. *The Journal of Pain, 12*(2), 228-235.

Schreurs, K. M. G., de Ridder, D. T. D., & Bensing, J. M. (2002). Fatigue in multiple sclerosis: Reciprocal relationships with physical disabilities and depression. *Journal of Psychosomatic Research, 53*(3), 775-781.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in Social Neuroscience* (pp. 17 - 44). New York, NY: Guilford Press.

Schulz, P., Kirschbaum, C., Pruessner, J., & Hellhammer, D. (1998). Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. *Stress Medicine, 14*(2), 91-97.

References

Schulz, P., Schlotz, W., & Becker, P. (2004). *Trierer Inventar zum Chronischen Stress (TICS) [Trier Inventory for Chronic Stress (TICS)]*. Gottingen: Hogrefe Verlag.

Schumann, E. M., Kümpfel, T., Then Bergh, F., Trenkwalder, C., Holsboer, F., & Auer, D. P. (2002). Activity of the hypothalamic-pituitary-adrenal axis in multiple sclerosis: correlations with gadolinium-enhancing lesions and ventricular volume. *Annals of Neurology, 51*(6), 763-767.

Schwartz, J. E., & Stone, A. A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology, 17*(1), 6-16.

Schwartz, J. E., & Stone, A. A. (2008). The analysis of real-time momentary data: a practical guide. In A. A. Stone, S. Shiffman, A. A. Atienza & l. Nebeling (Eds.), *The Science of Real-Time Data Capture*. New York, NY: Oxford University Press.

Schwartz, N. (2008). Retrospective and concurrent self-reports: the rationale for real-time data capture. In A. A. Stone, S. Shiffman, A. A. Atienza & L. Nebeling (Eds.), *The Science of Real-Time Data Capture*. New York, NY: Oxford University Press.

Schwartz, N. (2011). Why Researchers Should Think "Real-Time": A Cognitive Rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Press.

Schwid, S. R., Covington, M., Segal, B. M., & Goodman, A. D. (2002). Fatigue in multiple sclerosis: current understanding and future directions. *Journal of Rehabilitation Research and Development, 39*(2), 211-224.

Schwid, S. R., Petrie, M. D., Murray, R., Leitch, J., Bowen, J., Alquist, A.et al. (2003). A randomized controlled study of the acute and chronic effects of cooling therapy for MS. *Neurology, 60*(12), 1955-1960.

Sephton, S., & Spiegel, D. (2003). Circadian disruption in cancer: a neuroendocrine-immune pathway from stress to disease? *Brain, Behavior, and Immunity, 17*(5), 321-328.

Sfagos, C., Makis, A. C., Chaidos, A., Hatzimichael, E. C., Dalamaga, A., Kosma, K.et al. (2005). Serum ferritin, transferrin and soluble transferrin receptor levels in multiple sclerosis patients. *Multiple Sclerosis, 11*(3), 272-275.

Sharpe, M. C., Archard, L. C., Banatvala, J. E., Borysiewicz, L. K., Clare, A. W., David, A.et al. (1991). A report--chronic fatigue syndrome: guidelines for research. *Journal of the Royal Society of Medicine, 84*(2), 118-121.

Shen, J., Botly, L. C. P., Chung, S. A., Gibbs, A. L., Sabanadzovic, S., & Shapiro, C. M. (2006). Fatigue and shift work. *Journal of Sleep Research, 15*(1), 1-5.

Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology, 3*, 34.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1-32.

Shrout, P. E., & Lane, S. P. (2011). Psychometrics. In M. R. Mehl & T. A. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Press.

Sicotte, N. L., Kern, K. C., Giesser, B. S., Arshanapalli, A., Schultz, A., Montag, M.et al. (2008). Regional hippocampal atrophy in multiple sclerosis. *Brain, 131*(4), 1134-1141.

Siegert, R. J., & Abernethy, D. A. (2005). Depression in multiple sclerosis: a review. *Journal of Neurology, Neurosurgery & Psychiatry, 76*(4), 469-475.

SigmaPlot (Version 12.3) [Computer software]. San Jose, CA: Systat Software.

Simmons, R. D., Ponsonby, A.-L., van der Mei, I. A. F., & Sheridan, P. (2004). What affects your MS? Responses to an anonymous, Internet-based epidemiological survey. *Multiple Sclerosis, 10*(2), 202-211.

Simpson, S., Blizzard, L., Otahal, P., Van der Mei, I., & Taylor, B. (2011). Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry, 82*(10), 1132-1141.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence*. New York, NY: Oxford University Press.

Sjögren, E., Leanderson, P., & Kristenson, M. (2006). Diurnal saliva cortisol levels and relations to psychosocial factors in a population sample of middle-aged Swedish men and women. *International Journal of Behavioral Medicine, 13*(3), 193-200.

Skerrett, T. N., & Moss-Morris, R. (2006). Fatigue and social impairment in multiple sclerosis: The role of patients' cognitive and behavioral responses to their symptoms. *Journal of Psychosomatic Research, 61*(5), 587-593.

References

Smets, E. M. A., Garssen, B., Bonke, B., & De Haes, J. C. J. M. (1995). The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research, 39*(3), 315-325.

Smyth, J., Ockenfels, M. C., Porter, L., Kirschbaum, C., Hellhammer, D. H., & Stone, A. A. (1998). Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion. *Psychoneuroendocrinology, 23*(4), 353-370.

Smyth, N., Clow, A., Thorn, L., Hucklebridge, F., & Evans, P. (2013). Delays of 5-15 min between awakening and the start of saliva sampling matter in assessment of the cortisol awakening response. *Psychoneuroendocrinology, 38*(9), 1476-1483.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd Edition ed.). London, UK: SAGE Publications.

Sohl, S. J., & Friedberg, F. (2008). Memory for fatigue in chronic fatigue syndrome: relationships to fatigue variability, catastrophizing, and negative affect. *Behavioral Medicine, 34*(1), 29-38.

Sonnenschein, M., Sorbi, M. J., van Doornen, L. J. P., & Maas, C. J. M. (2006). Feasibility of an electronic diary in clinical burnout. *International Journal of Behavioral Medicine, 13*(4), 315-319.

Sospedra, M., & Martin, R. (2005). Immunology of multiple sclerosis. *Annual Review of Immunology, 23*(1), 683-747.

Sriram, S., & Steiner, I. (2005). Experimental allergic encephalomyelitis: A misleading model of multiple sclerosis. *Annals of Neurology, 58*(6), 939-945.

Stankoff, B. M. D. P., Waubant, E. M. D. P., Confavreux, C., Edan, G., Debouverie, M., Rumbach, L.et al. (2005). Modafinil for fatigue in MS A randomized placebo-controlled double-blind study. *Neurology, 64*(7), 1139-1143.

Stazzone, L., & Brown, B. (2012). Symptom management. In H. L. Weiner & J. M. Stankiewicz (Eds.), *Multiple Sclerosis: Diagnosis and Treatment*. Chichester, UK: Wiley-Blackwell.

Steensberg, A., Fischer, C. P., Keller, C., Møller, K., & Pedersen, B. K. (2003). IL-6 enhances plasma IL-1ra, IL-10, and cortisol in humans. *American Journal of Physiology - Endocrinology And Metabolism, 285*(2), E433-E437.

Steinman, L., & Zamvil, S. S. (2006). How to successfully apply animal studies in experimental allergic encephalomyelitis to research on multiple sclerosis. *Annals of Neurology, 60*(1), 12-21.

Sterling, P., & Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In S. Fisher & J. Reason (Eds.), *Handbook of Life Stress, Cognition, and Health* (pp. 629-649). Chichester, UK: John Riley.

Sternberg, E. M. (2006). Neural regulation of innate immunity: a coordinated nonspecific host response to pathogens. *Nature Reviews Immunology, 6*(4), 318-328.

Stetler, C., Dickerson, S. S., & Miller, G. E. (2004). Uncoupling of social zeitgebers and diurnal cortisol secretion in clinical depression. *Psychoneuroendocrinology, 29*(10), 1250-1259.

Stetler, C., & Miller, G. E. (2011). Depression and hypothalamic-pituitary-adrenal activation: A quantitative summary of four decades of research. *Psychosomatic Medicine, 73*(2), 114-126.

Stone, A. A., Broderick, J. E., Kaell, A. T., DelesPaul, P. A. E. G., & Porter, L. E. (2000). Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics? *The Journal of Pain, 1*(3), 212-217.

Stone, A. A., Broderick, J. E., Porter, L. S., & Kaell, A. T. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis & Rheumatism, 10*(3), 185-193.

Stone, A. A., Broderick, J. E., Schneider, S., & Schwartz, J. E. (2012). Expanding options for developing outcome measures from momentary assessment data. *Psychosomatic Medicine, 74*(4), 387-397.

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain, 104*(1–2), 343-351.

References

Stone, A. A., Broderick, J. E., Shiffman, S. S., & Schwartz, J. E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 107*(1–2), 61-69.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16*(3), 199-202.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*(3), 236-243.

Strachan, D. P. (1989). Hay fever, hygiene, and household size. *BMJ: British Medical Journal, 299*(6710), 1259-1260.

Strachan, D. P. (2000). Family size, infection, and atopy: the first decade of the "hygiene hypothesis". *Thorax, 55*(Suppl 1), S2-S10.

Strickland, P., Morriss, R., Wearden, A., & Deakin, B. (1998). A comparison of salivary cortisol in chronic fatigue syndrome, community depression and healthy controls. *Journal of Affective Disorders, 47*(1-3), 191-194.

Strober, L. B., & Arnett, P. A. (2005). An examination of four models predicting fatigue in multiple sclerosis. *Archives of Clinical Neuropsychology, 20*(5), 631-646.

Stuifbergen, A. K., & Rogers, S. (1997). The experience of fatigue and strategies of self-care among persons with multiple sclerosis. *Applied Nursing Research, 10*(1), 2-10.

Sudhaus, S., Fricke, B., Stachon, A., Schneider, S., Klein, H., von During, M.et al. (2009). Salivary cortisol and psychological mechanisms in patients with acute versus chronic low back pain. *Psychoneuroendocrinology, 34*(4), 513-522.

Sullivan, M. J., Weinshenker, B., Mikail, S., & Bishop, S. R. (1995). Screening for major depression in the early stages of multiple sclerosis. *The Canadian Journal of Neurological Sciences, 22*(3), 228-231.

Swain, M. G. (2000). Fatigue in chronic disease. *Clinical Science 99*(1), 1-8.

Tachibana, N., Howard, R. S., Hirsch, N. P., Miller, D. H., Moseley, I. F., & Fish, D. (1994). Sleep problems in multiple sclerosis. *European Neurology, 34*(6), 320-323.

Tak, L. M., Cleare, A. J., Ormel, J., Manoharan, A., Kok, I. C., Wessely, S.et al. (2011). Meta-analysis and meta-regression of hypothalamic-pituitary-adrenal axis activity in functional somatic disorders. *Biological Psychology, 87*(2), 183-194.

Tedeschi, G., Dinacci, D., Lavorgna, L., Prinster, A., Savettieri, G., Quattrone, A.et al. (2007). Correlation between fatigue and brain atrophy and lesion load in multiple sclerosis patients independent of disability. *Journal of the Neurological Sciences, 263*(1–2), 15-19.

Téllez, N., Río, J., Tintoré, M., Nos, C., Galán, I., & Montalban, X. (2006). Fatigue in multiple sclerosis persists over time: a longitudinal study. *Journal of Neurology, 253*(11), 1466-1470.

Ten, S., New, M., & Maclaren, N. (2001). Addison's disease 2001. *Journal of Clinical Endocrinology and Metabolism, 86*(7), 2909-2922.

Then-Bergh, F., Kümpfel, T., Trenkwalder, C., Rupprecht, R., & Holsboer, F. (1999). Dysregulation of the hypothalamo-pituitary-adrenal axis is related to the clinical course of MS. *Neurology, 53*(4), 772-777.

Thorn, L., Hucklebridge, F., Esgate, A., Evans, P., & Clow, A. (2004). The effect of dawn simulation on the cortisol response to awakening in healthy participants. *Psychoneuroendocrinology, 29*(7), 925-930.

Thorn, L., Hucklebridge, F., Evans, P., & Clow, A. (2006). Suspected non-adherence and weekend versus week day differences in the awakening cortisol response. *Psychoneuroendocrinology, 31*(8), 1009-1018.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Research, 34*(4), 414-424.

Torres-Harding, S., & Jason, L. A. (2005). What is Fatigue? History and Epidemiology. In J. DeLuca (Ed.), *Fatigue As a Window to the Brain*. Cambridge, UK: The MIT Press.

Torres-Harding, S., Sorenson, M., Jason, L., Reynolds, N., Brown, M., Maher, K.et al. (2008). The associations between basal salivary cortisol and illness symptomatology in chronic fatigue syndrome. *Journal of Applied Biobehavioral Research, 13*(3), 157-180.

Tremlett, H., Luscombe, D. K., & Wiles, C. M. (1998). Use of corticosteroids in multiple sclerosis by consultant neurologists in the United Kingdom. *Journal of Neurology, Neurosurgery & Psychiatry, 65*(3), 362-365.

References

Tremlett, H., Zhao, Y., Joseph, J., & Devonshire, V. (2008). Relapses in multiple sclerosis are age- and time-dependent. *Journal of Neurology, Neurosurgery & Psychiatry, 79*(12), 1368-1374.

Tremlett, H., Zhao, Y., Rieckmann, P., & Hutchinson, M. (2010). New perspectives in the natural history of multiple sclerosis. *Neurology, 74*(24), 2004-2015.

Trojan, D. A., Arnold, D., Collet, J.-P., Shapiro, S., Bar-Or, A., Robinson, A.et al. (2007). Fatigue in multiple sclerosis: association with disease-related, behavioural and psychosocial factors. *Multiple Sclerosis, 13*(8), 985-995.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9*(1), 151-176.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? *Journal of Educational and Behavioral Statistics, 35*(2), 215-247.

Van den Brink, M., Bandell-Hoekstra, E. N. G., & Abu-Saad, H. H. (2001). The occurrence of recall bias in pediatric headache: A comparison of questionnaire and diary data. *Headache: The Journal of Head and Face Pain, 41*(1), 11-20.

Van den Broeck, A., De Cuyper, N., De Witte, H., & Vansteenkiste, M. (2010). Not all job demands are equal: Differentiating job hindrances and job challenges in the Job Demands–Resources model. *European Journal of Work and Organizational Psychology, 19*(6), 735-759.

van Eck, M., Berkhof, H., Nicolson, N., & Sulon, J. (1996). The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. *Psychosomatic Medicine, 58*(5), 447-458.

van Kessel, K., & Moss-Morris, R. (2006). Understanding multiple sclerosis fatigue: A synthesis of biological and psychological factors. *Journal of Psychosomatic Research, 61*(5), 583-585.

van Kessel, K., Moss-Morris, R., Willoughby, E., Chalder, T., Johnson, M. H., & Robinson, E. (2008). A randomized controlled trial of cognitive behavior therapy for multiple sclerosis fatigue. *Psychosomatic Medicine, 70*(2), 205-213.

van't Leven, M., Zielhuis, G. A., van der Meer, J. W., Verbeek, A. L., & Bleijenberg, G. (2010). Fatigue and chronic fatigue syndrome-like complaints in the general population. *European Journal of Public Health, 20*(3), 251-257.

Vaucher, P., Druais, P.-L., Waldvogel, S., & Favrat, B. (2012). Effect of iron supplementation on fatigue in nonanemic menstruating women with low ferritin: a randomized controlled trial. *Canadian Medical Association Journal, 184*(11), 1247-1254.

Veauthier, C., Radbruch, H., Gaede, G., Pfueller, C. F., Dörr, J., Bellmann-Strobl, J.et al. (2011). Fatigue in multiple sclerosis is closely related to sleep disorders: A polysomnographic cross-sectional study. *Multiple Sclerosis Journal, 17*(5), 613-622.

Veen, G., van Vliet, I. M., DeRijk, R. H., Giltay, E. J., van Pelt, J., & Zitman, F. G. (2011). Basal cortisol levels in relation to dimensions and DSM-IV categories of depression and anxiety. *Psychiatry Research, 185*(1), 121-128.

Verdon, F., Burnand, B., Fallab Stubi, C.-L., Bonard, C., Graff, M., Michaud, A.et al. (2003). Iron supplementation for unexplained fatigue in non-anaemic women: double blind randomised placebo controlled trial. *British Medical Journal, 326*(7399), 1124.

Vickers, A. J., & Altman, D. G. (2001). Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ: British Medical Journal, 323*(7321), 1123-1124.

von Baeyer, C. L. (1994). Reactive effects of measurement of pain. *The Clinical Journal Of Pain, 10*(1), 18-21.

von Bardeleben, U., & Holsboer, F. (1989). Cortisol response to a combined dexamethasone-human corticotrophin-releasing hormone challenge in patients with depression. *Journal of Neuroendocrinology, 1*(6), 485-488.

Vucic, S., Burke, D., & Kiernan, M. C. (2010). Fatigue in multiple sclerosis: Mechanisms and management. *Clinical Neurophysiology, 121*(6), 809-817.

Walley, T., & Mantgani, A. (1997). The UK general practice research database. *The Lancet, 350*(9084), 1097-1099.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*(6), 473-483.

References

Wei, T., & Lightman, S. L. (1997). The neuroendocrine axis in patients with multiple sclerosis. *Brain, 120*(6), 1067-1076.

Wessely, S. (2001). Chronic fatigue: Symptom and syndrome. *Annals of Internal Medicine, 134*(9 Part 2), 838-843.

Wessely, S., Chalder, T., Hirsch, S., Wallace, P., & Wright, D. (1996). Psychological symptoms, somatic symptoms, and psychiatric disorder in chronic fatigue and chronic fatigue syndrome: a prospective study in the primary care setting. *American Journal of Psychiatry, 153*(8), 1050-1059.

White, K., Speechley, M., Harth, M., & Ostbye, T. (2000). Co-existence of chronic fatigue syndrome with fibromyalgia syndrome in the general population: A controlled study. *Scandinavian Journal of Rheumatology, 29*(1), 44-51.

Whitehead, L. (2009). The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. *Journal of Pain and Symptom Management, 37*(1), 107-128.

Whitworth, J. A., Williamson, P. M., Mangos, G., & Kelly, J. J. (2005). Cardiovascular consequences of cortisol excess. *Journal of Vascular Health and Risk Management, 1*(4), 291-299.

Wilhelm, I., Born, J., Kudielka, B. M., Schlotz, W., & Wüst, S. (2007). Is the cortisol awakening rise a response to awakening? *Psychoneuroendocrinology, 32*(4), 358-366.

Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D., Ebers, G. C., & null. (2003). Twin concordance and sibling recurrence rates in multiple sclerosis. *Proceedings of the National Academy of Sciences of the United States of America, 100*(22), 12877-12882.

Williams, G. H., & Dluhy, R. G. (2006). Disorders of the adrenal cortex. In J. L. Jameson (Ed.), *Harrison's Endocrinology* (Vol. 323, pp. 113-150). Chicago, IL: McGraw-Hill.

Wolfe, F., Hawley, D. J., & Wilson, K. (1996). The prevalence and meaning of fatigue in rheumatic disease. *The Journal of Rheumatology, 23*(8), 1407-1417.

Wood, B., van der Mei, I. A. F., Ponsonby, A.-L., Pittas, F., Quinn, S., Dwyer, T.et al. (2012). Prevalence and concurrence of anxiety, depression and fatigue over time in multiple sclerosis. *Multiple Sclerosis Journal, 19*(2), 217-224.

Wood, B., Wessely, S., Papadopoulos, A., Poon, L., & Checkley, S. (1998). Salivary cortisol profiles in chronic fatigue syndrome. *Neuropsychobiology, 37*, 1-4.

Wu, S. M., & Amtmann, D. (2013). Psychometric evaluation of the Perceived Stress Scale in multiple sclerosis. *ISRN Rehabilitation, Article ID 608356*, 1-9.

Wüst, S., Wolf, J., Hellhammer, D. H., Federenko, I., Schommer, N., & Kirschbaum, C. (2000). The cortisol awakening response - normal values and confounds. *Noise and Health, 2*(7), 79-88.

Yaldizli, Ö., Glassl, S., Sturm, D., Papadopoulou, A., Gass, A., Tettenborn, B.et al. (2011). Fatigue and progression of corpus callosum atrophy in multiple sclerosis. *Journal of Neurology, 258*(12), 2199-2205.

Yang, C., & Hsu, Y. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors, 10*(8), 7772-7788.

Yanovski, J. A., Cutler, G. B., Jr., Chrousos, G. P., & Nieman, L. K. (1998). The dexamethasone-suppressed corticotropin-releasing hormone stimulation test differentiates mild Cushing's disease from normal physiology. *The Journal of Clinical Endocrinology & Metabolism, 83*(2), 348-352.

Young, A. H., Sharpe, M., Clements, A., Dowling, B., Hawton, K. E., & Cowen, P. J. (1998). Basal activity of the hypothalamic-pituitary-adrenal axis in patients with the chronic fatigue syndrome (neurasthenia). *Biological Psychiatry, 43*(3), 236-237.

Young, E. A., Abelson, J., & Lightman, S. L. (2004). Cortisol pulsatility and its role in stress regulation and health. *Frontiers in Neuroendocrinology, 25*(2), 69-76.

Young, J. S., Bourgeois, J. A., Hilty, D. M., & Hardin, K. A. (2008). Sleep in hospitalized medical patients, Part 1: Factors affecting sleep. *Journal of Hospital Medicine, 3*(6), 473-482.

Ysrraelit, M. C., Gaitán, M. I., Lopez, A. S., & Correale, J. (2008). Impaired hypothalamic-pituitary-adrenal axis activity in patients with multiple sclerosis. *Neurology, 71*(24), 1948-1954.

Ziemssen, T. (2009). Multiple sclerosis beyond EDSS: depression and fatigue. *Journal of the Neurological Sciences, 277, Supplement 1*(0), S37-S41.

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.